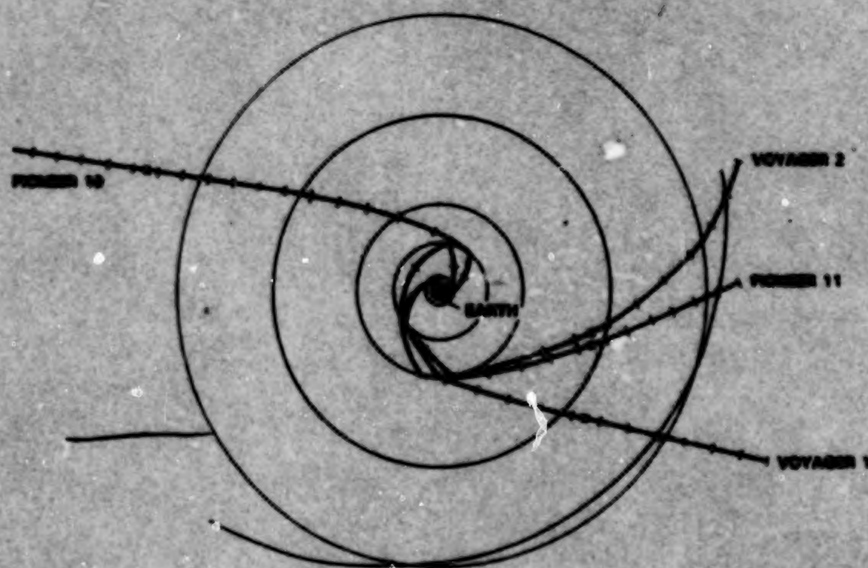
NASA Conference Publication 2464

Essays in Space Science



Proceedings of a symposium held at
NASA Goddard Space Flight Center
Greenbelt, Maryland
April 23, 1985

NASA

NASA Conference Publication 2464

Essays in Space Science

Edited by
R. Ramaty, T. L. Cline, and J. F. Ormes
NASA Goddard Space Flight Center
Greenbelt, Maryland

Proceedings of a symposium held at
NASA Goddard Space Flight Center
Greenbelt, Maryland
April 23, 1985

NASA
National Aeronautics
and Space Administration

Scientific and Technical
Information Branch

1987

Library of Congress Cataloging-in-Publication Data

Essays in space science.

(NASA conference publication ; 2464)
1. Space sciences--Congresses. 2. Astrophysics--Congresses. 3. Cosmic rays--Congresses. 4. Astronomy--Congresses. 5. McDonald, Frank B. (Frank Bethune)
I. McDonald, Frank B. (Frank Bethune) II. Ramaty, R.
III. Cline, T. L. IV. Ormes, J. F., 1939- .
V. Series.
Qb495.E83 1987 500.5 87-7504


JAMES A. VAN ALLEN
AND
FRANK McDONALD
13 MAR., 1956
IOWA CITY, IOWA

PREFACE

A Space Science Symposium was held at the Goddard Space Flight Center (GSFC) in Greenbelt, Maryland on April 23, 1985. The topics discussed covered a broad segment of space research and were an acknowledgement of Dr. Frank B. McDonald's personal involvement in many of these efforts.

Drs. James Van Allen (University of Iowa), Miriam Forman (Stony Brook), Eugene Parker (University of Chicago), and Kinsey Anderson (University of California, Berkeley) made presentations in the morning, and Drs. Norman Ness (GSFC), Catherine Cesarsky (Saclay, France), William Webber (University of New Hampshire), Allan Jacobson (Jet Propulsion Laboratory), Riccardo Giacconi (Space Telescope Science Institute), and John Naugle (Fairchild Corporation) in the afternoon. Introductory remarks by Drs. Noel Hinners (Director, GSFC) and John Simpson (University of Chicago) and an after-dinner speech by Dr. Phyllis Freier (University of Minnesota) bracketed the agenda. All of these texts are included in this volume. Additional manuscripts by Frank's colleagues which could not be fitted as talks into the 1-day schedule are also included. These papers were authored or co-authored by V. K. Balasubrahmanyan (GSFC), W. R. Binns (Washington University), E. A. Boldt (GSFC), T. L. Cline (GSFC), W. M. Dougherty (University of California, Berkeley), D. S. Evans (NOAA), C. E. Fichtel (GSFC), T. L. Garrard (Caltech), M. G. Hauser (GSFC), M. H. Israel (Washington University), J. Klarmann (Washington University), S. S. Holt (GSFC), J. F. Ormes (GSFC), R. Ramaty (GSFC), E. C. Stone (Caltech), R. E. Streitmatter (GSFC) and C. J. Waddington (University of Minnesota).

All of the speakers and authors had been closely associated with Frank McDonald during various aspects of his career. The totality of the manuscripts forming "Essays in Space Science" were chosen so as to sample the scientific areas influenced by him in a significant manner. These texts have been arranged into the three broad areas: particles and fields of the solar systems, cosmic ray astrophysics, and gamma ray, X-ray and infrared astronomies. Tributes to F. B. McDonald and a history of space science form the Afterword.

We wish to thank Mr. Robert Frey for his meticulous scrutiny and many efforts with the details of the preparation of the text for publication and Mrs. Evelyn Schronce for her help with the Symposium organization and her patient assistance during all phases of the creation of this book.

R. Ramaty
T. L. Cline
J. F. Ormes

CONTENTS

IV. AFTERWORD

PART I:

PARTICLES AND FIELDS OF THE SOLAR SYSTEM

N87

24248

UNCLAS

N87-24248

1

MAGNETOSPHERES OF THE OUTER PLANETS

James A. Van Allen
Department of Physics and Astronomy
University of Iowa
Iowa City, Iowa 52242

1. INTRODUCTION

Anyone who wishes to become properly grounded in magnetospheric physics is well advised to start with three great monographs: Chapman and Bartel's *Geomagnetism* [1940], Alfvén's *Cosmical Electrodynamics* [1950], and Störmer's *The Polar Aurora* [1955].

The subject has acquired its present vigor and broad participation only during the 27 years since 1958, when we found that enormous numbers of energetic charged particles are durably trapped in Earth's external magnetic field.

The magnetosphere of Earth is the prototypical planetary magnetosphere. It has been investigated intensively and may be said to be understood to "first order", though many of its details continue to be baffling and controversial.

Meanwhile, we have been proceeding with the investigation of particle and field phenomena associated with other planetary bodies—the Moon, Mercury, Venus, Mars, Jupiter, and Saturn. There are significant plasma physical effects at the Moon and at Mercury, Venus, and Mars, but only Jupiter and Saturn join Earth in exhibiting fully developed magnetospheres.

The three latter cases have a certain gross similarity but each is distinctively different in detail, thus giving rise to Frank McDonald's [1980] famous remark: "If you've seen one magnetosphere, you haven't seen them all." It is reasonable

to expect that the prospective Voyager 2 investigations of Uranus in 1986 and Neptune in 1989 will add further support to this remark.

Because of the availability of a massive body of literature including several book-length monographs and review papers on the magnetospheres of Earth, Jupiter, and Saturn, I decided not to attempt a twenty-minute digest of this knowledge but rather to review some general considerations of an elementary nature and to present some speculations.

2. CONDITIONS FOR THE EXISTENCE OF A PLANETARY MAGNETOSPHERE

A common statement is that a planet will have a magnetosphere if and only if it: (a) is "sufficiently strongly magnetized" and (b) is subjected to the flow of the solar wind.

Such a statement contains a certain measure of validity but requires further scrutiny.

On the subject of planetary magnetism, the following section is adapted from a paper that I wrote in 1976 [Van Allen, 1977].

There are five qualitatively different types of magnetism that a planetary body can exhibit:

(a) Remanent ferromagnetism in cool crustal material.

(b) Electromagnetism caused by currents in an electrically conductive interior, such currents being driven by self-excited dynamo electromotive forces generated by the convective flow of material. This mechanism requires a hot fluid interior and planetary rotation at a "sufficiently rapid rate".

(c) Electromagnetism of type (b) at some remotely previous epoch, with subsequent resistive-inductive decay of the current systems after the electromotive forces have become negligible.

(d) Electromagnetism caused by systems of currents induced in the conducting ionosphere of the planet by fluctuating magnetic fields in the solar wind and/or driven by motional electromotive forces caused by the relative motion of magnetic fields in the solar wind. In either case, the electrical circuit may be closed in part through the conductive interplanetary medium.

(e) Electromagnetism similar to type (d), but with the induced currents in conducting portions of the planetary body itself.

Most of the interiors of the above mentioned seven celestial bodies (with the possible exception of the Moon) are thought to be at temperatures above the Curie temperature of ferromagnetic materials ($\approx$ 1000 K) if, indeed, such materials are present; hence, remanent ferromagnetism, if any, must be confined to the outer crust of the bodies. For a large, rotating planet having a fluid interior, there is no theory of type (b) magnetism that proceeds from first principles to a confident quantitative prediction of the magnetic properties of the planet.

In order that a planetary body have a magnetosphere of durably trapped particles, it is necessary that its dipole moment be sufficiently great that there are closed magnetic shells such that particles can drift in longitude without striking the body (or its appreciable atmosphere) or without escaping from the system.

In the vacuum case, the magnetic field extends to infinity and the criterion for durable trapping of a single test particle is derivable from Störmer theory. The total population of trapped, non-interacting particles is limited by the further criterion that the volume density of kinetic energy of charged particles is less than or of the order of $B^2/8\pi$, where B is the local magnetic field strength. A realistic physical source of particles is, of course, cosmic ray albedo neutron decay. I consider that a full solution of the self-consistent vacuum case would be a worthy theoretical exercise but, to my knowledge, no one has produced such a solution. The lack of interest in the vacuum case stems from the fact that flowing plasma, or at least plasma, appears to be ubiquitous throughout the universe.

In the presence of flowing plasma, the approximate criterion for the existence of a magnetosphere is that the magnetohydrodynamic stagnation distance r on the upstream side of the planet exceeds the radius of its surface or appreciable atmosphere. In the case that plasma within the magnetosphere exerts a negligible pressure, the magnitude of r is given by

$$nmv^2 = \frac{M^2}{2\pi r^6} \qquad (1)$$

wherein n, m, and v are the number density, atomic mass, and relative bulk velocity of the ions in the plasma and M is the body's magnetic dipole moment. The plasma in question may be the solar wind out to the heliopause or the interstellar wind beyond the heliopause. For planetary satellites within a planet's magnetosphere, the flowing plasma may be that co-rotating with the planet.

Inside the heliosphere, v for the solar wind is independent of the distance R from the sun and n is inversely proportional to R^2. Hence by Equation (1)

$$\frac{r_P}{r_E} = \left(\frac{M_P\ R_P}{M_E\ R_E}\right)^{1/3} \qquad (2)$$

where the subscripts E and P refer to Earth and to any other planet, respectively.

It is seen from Equation (2) that Earth would have the same size magnetosphere as it now does if it were at a heliocentric distance of 50 AU and its magnetic moment were reduced by a factor of 50. This statement assumes, of course, that the heliopause lies beyond 50 AU, as now seems likely.

By Equation (1), it is noted that if Earth were placed outside the heliopause in the nearby interstellar medium ($n \sim 0.05\ cm^{-3}$, $v \sim 20\ km\ s^{-1}$), r would be unchanged if the planet's magnetic moment were only 1/200 of its present value. This example illustrates the fact that the solar wind is not essential for producing a magnetosphere.

A further case of general interest contemplates a magnetized planet immersed in a stationary plasma. If the planet were not rotating, it would be simply a large Langmuir probe with no magnetospheric properties. But if the planet is rotating, however slowly, there is a corresponding unipolar electric field and the nearby plasma will co-rotate with the planet out to the radius at which the co-rotational speed is equal to the Alfvén speed. At the outer boundary of the co-rotating plasma, there are presumably instability effects that result in the generation of waves and acceleration of particles. Hence, even in the case of nonflowing plasma a magnetosphere will exist. Cases of this nature are treated in the theory of pulsars, though for rotational rates far greater than those of planets.

The foregoing remarks suggest the great variety of magnetospheres that may, and probably do, exist.

3. SOURCE OF MAGNETOSPHERIC PARTICLES

Potential sources of energetic particles in a planet's magnetosphere are as follows:

(a) The solar wind

(b) Solar energetic particles

(c) Primary cosmic rays

(d) Secondary particles from cosmic-ray interactions in the planet's atmosphere, rings, and satellites

(e) Ionized gas from the planet's ionosphere

(f) Gas sputtered from rings and satellites by particle and photon bombardment

(g) Gas emitted volcanically or outgassed from rings and satellites

Plasma physical phenomena associated with the Moon, Mercury, Venus, and Mars are attributed principally to particles from source (a), with perhaps an admixture of particles from source (e).

The quasi-thermal plasma and low energy particles within Earth's magnetosphere are also primarily from source (a) and secondarily from source (e), as judged by elemental composition and energy spectra. However, higher energy ($E \gtrsim 0.2$ MeV) particles come primarily from source (d), with some admixture of particles from source (b).

The quasi-thermal plasma and low energy particles in Jupiter's magnetosphere are identified as dominantly from source (g), the volcanically active satellite Io being the principal contributor, but there are also (probably) significant contributions from sources (a), (e), and (f). Sources (b) and (d) of very energetic particles are presumed to be operative but particles from these sources have not been identified conclusively.

The various potential sources of particles in Saturn's magnetosphere have been assessed as follows: Very energetic protons ($E_p \gtrsim 10$ MeV) are from source (d). Protons having $E_p \lesssim 1$ MeV and electrons having $0.035 < E_e <$ a few MeV are principally from the solar wind, source (a), and to a lesser extent from source (b). Electrons and protons of lesser energies are apparently from sources (e), (f), and (g) as well as from the ionosphere of Titan.

4. IN SITU ENERGIZATION OF PARTICLES

Particles that are injected into a magnetosphere, from whatever source and in whatever manner, are energized and diffused spatially by fluctuating magnetic and electric fields (including those in plasma waves) and convected and energized by quasi-steady electric fields. These very complex processes are essential to the overall character of the particle population but they also tend to confuse the process of identification of the sources and sinks of the particles.

Irrespective of detailed processes for the generation of plasma waves and electromagnetic radiation and for the acceleration and diffusion of particles, it

appears that all such magnetospheric processes derive their power from three basic sources:

(a) The kinetic energy of flowing plasma (e.g., the solar wind)

(b) The rotational energy of the planet

(c) The orbital energy of satellites

Sunlight contributes to establishing the conditions for energy transfer by ionizing atmospheric gases and by photon sputtering and ionization of solid surface material but apparently contributes little to gross energetics.

The coupling between any one of the three sources of energy and the particle population must be such as to generate electric fields since, apart from gravitational fields, acceleration of a charged particle can be accomplished only by an electric field.

The power flux of the solar wind at 1 AU is typically 0.4 erg $(cm^2 s)^{-1}$. The cross-sectional area of Earth's magnetosphere perpendicular to the solar wind flow is that of a circle of approximate radius 14 planetary radii or 2.5×10^{20} cm^2. Hence, the total power that is potentially available from the solar wind is of the order of 1×10^{20} erg s^{-1}, on an average day. During days of high solar activity the power flux increases by as much as an order of magnitude.

The solar wind power exceeds that required for all dissipative processes in Earth's magnetosphere by a factor of the order of 100. The coupling of this power into the magnetosphere apparently occurs by way of the motional electromotive force induced in the magnetosheath.

The total rotational kinetic energy of Earth is 2×10^{36} erg, an enormous amount relative to all magnetospheric requirements. The coupling here is exhibited by co-rotation of plasma and the maintenance of the corresponding system of electrical currents in the plasma and in the ionosphere. The best present estimates are that the power extracted from Earth's rotational energy is much less than that extracted from the solar wind flow.

At Jupiter, the situation is the reverse of that at Earth, and it appears that most of the power for magnetospheric processes comes from the rotational energy of the planet and the orbital energy of the innermost Galilean satellite, Io. The gas emitted volcanically from Io plays a central role in Jupiter's magnetosphere. Nonetheless, the long magnetotail (several AU in length) of Jupiter certifies the importance of the solar wind flow in establishing the topology of its outer magnetosphere.

Saturn also exhibits a long magnetotail but the energetics of its magnetosphere lie between those of Earth and Jupiter, with perhaps comparable contributions from the solar wind and the planet's rotation. Titan with its dense atmosphere lies in the outer fringes of Saturn's magnetosphere and hence has a less significant role in the physics of Saturn's magnetosphere than does Io in Jupiter's.

5. THE MAGNETOSPHERE OF URANUS

One of the most exciting near-term prospects for magnetospheric physicists is the encounter of Voyager 2 with Uranus during the period around closest approach on January 24, 1986.

It has been established by Pioneer 10 that the radial flow of the solar wind extends to and far beyond the orbits of Uranus and also Neptune [Barnes and Gazis, 1984].

There is, as yet, little quantitative evidence on the magnetic moment of Uranus. However, it is reasonable to expect that the value of its moment is of the order of 4×10^{27} gauss cm^3 or about 0.2 gauss a_U^3 where a_U is the planet's equatorial radius.

The basis for this empirical expectation is shown in Table I and Figure 1, which summarize existing knowledge of the magnetic moments of planets as a function of their rotational angular momenta. The two line segments labeled Uranus and Neptune are drawn vertically at the approximately known values of their respective angular momenta and the lengths of the segments

Table I

Angular Momenta and Magnetic Moments of Planets

	$I\omega$	M	$\frac{M}{I\omega} \times 10^{15}$
	$g\ cm^2\ s^{-1}$	$gauss\ cm^3$	$gauss\ cm\ s\ g^{-1}$
Mercury	9.74 E 36	2.4 E 22	2.5
Venus	1.82 E 38	< 3 E 21	< 0.02
Earth	5.859 E 40	7.92 E 25	1.35
Mars	1.98 E 39	1.4 E 22	0.007
Jupiter	4.19 E 45	1.53 E 30	0.37
Saturn	7.03 E 44	4.32 E 28	0.061
Uranus	1.52 E 43	—	—
Neptune	2.07 E 43	—	—
Pluto	~ 3 E 36	—	—
Moon	2.36 E 36	< 4 E 20	< 0.2

Note: a E b = $a \times 10^b$

4/17/84

suggest ranges of values of their magnetic moments within which actual values will not be astonishing.

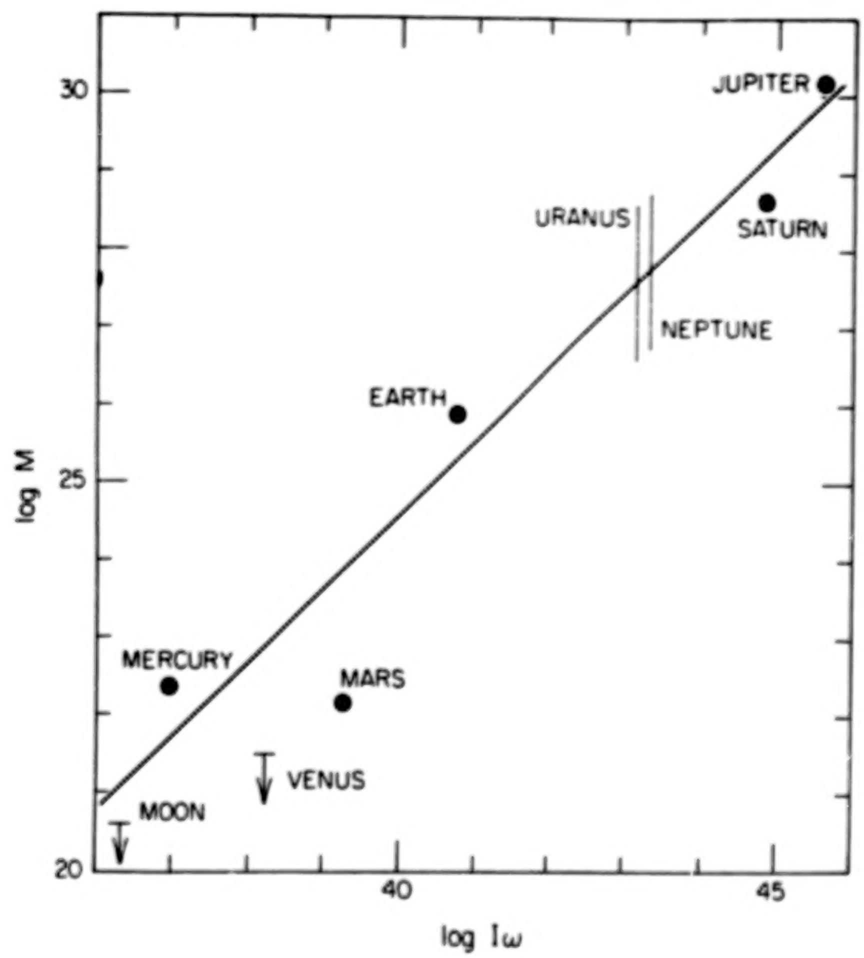


Figure 1. The empirical relationship between magnetic dipole moments M of planets and their rotational angular momenta Iω. The line segments labeled Uranus and Neptune are drawn vertically at the approximately known magnitudes of their angular momenta. The lengths of the segments span ranges of M within which it would not be astonishing to find their actual values.

Based on Equation (1) and the above guess as to the magnitude of the magnetic moment of Uranus, one calculates a stand-off distance of 26 planetary radii, a value that is proportional to the cube root of the assumed moment.

Within this radial distance from the center of the planet, it is reasonable to expect a well-developed magnetosphere, albeit one of extraordinary properties because of the approximately axial alignment of the rotational axis with the planet-sun line in 1986 as shown in Figure 2 [Van Allen, 1977]. If the magnetic axis of the planet is approximately parallel to its rotational axis, then the co-rotational equipotential surfaces are turned by 90° relative to the

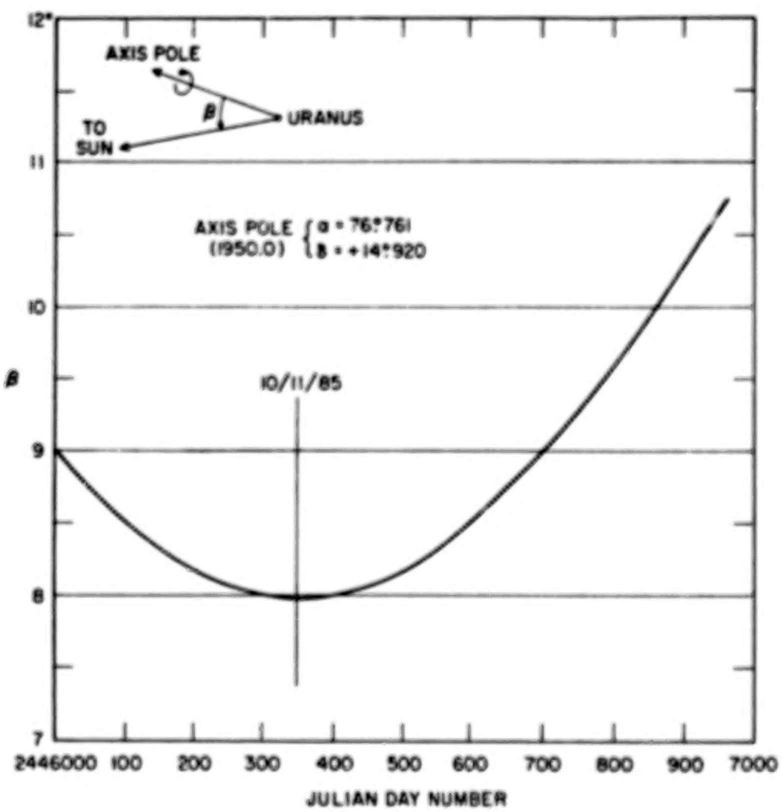
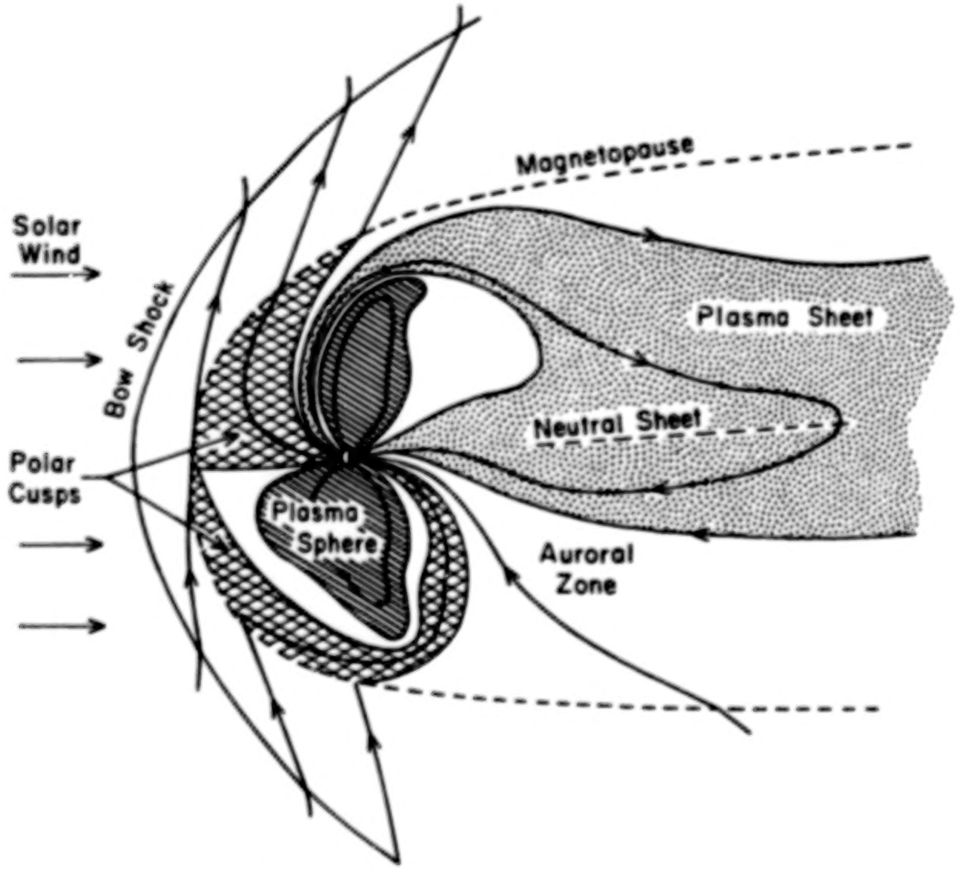


Figure 2. The time dependence of the angle β between the rotational axis of Uranus and the planet-sun line [Van Allen, 1977].

transverse potential surfaces that are attributed to solar wind flow—as compared to the situations at Earth, Jupiter, and Saturn. This case has been discussed in a preliminary way by Siscoe [1971, 1975] (Figure 3) and by Olson. If the magnetic axis is inclined markedly to the rotational axis, an even more exotic magnetosphere may be expected because of the large dirunal variation that will occur in this case.



HYPOTHETICAL MAGNETOSPHERE OF URANUS
(AFTER SISCOE)

Figure 3. Hypothetical topology of a Uranian magnetosphere during the epoch of pole-on presentation to the solar wind [Siscoe, 1975].

Recent observations by the International Ultraviolet Explorer of auroral optical emissions from Uranus provide the principal observational evidence thus far for the existence of a well-developed magnetosphere [Durrance and Moos, 1982; Clarke, 1982; and Caldwell, Wagener, and Owen, 1983]. Less direct evidence depends on the suggestion of Chang and Lanzerotti [1978] that the low optical albedos of the satellites and rings of Uranus are the result of trapped particle bombardment and consequent carbonization of methane ice on their surfaces.

An important feature of the Uranian magnetosphere is the presence therein of five satellites, in close regular orbits (Table II) [Dermott, 1984], and nine thin rings which lie between 1.59 and 1.96 planetary radii [Elliot, 1984]. As at Jupiter and Saturn, these elements of the Uranian system doubtless have profound effects on the absorption and possibly the emission and acceleration of charged particles.

Table II

Satellites of Uranus

Satellite	Orbital Radius (a_U)	Body Radius (km)
V Miranda	5.0	220
I Ariel	7.3	660
II Umbriel	10.2	560
III Titania	16.7	800
IV Oberon	21.7	815

During the Voyager 2 encounter with Uranus, Pioneer 11 will be relatively nearby as shown by Figure 4. Hence, Pioneer 11 will be able to provide valuable

observations of the solar wind, the magnetic field, and energetic particle intensity in the nearby interplanetary medium before, during, and after that encounter [Van Allen, 1984]. All of these quantities are significant in determining the state of Uranus' magnetosphere and fluctuations thereof.

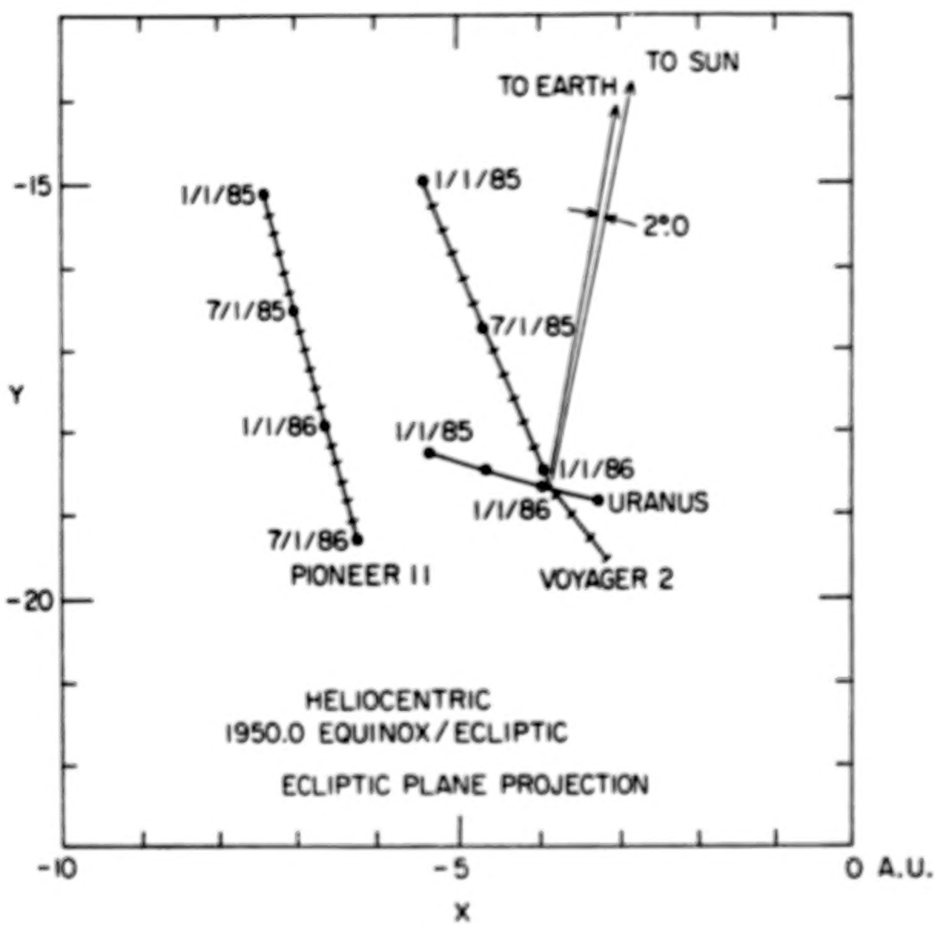


Figure 4. Ecliptic plane projection of the trajectories of Pioneer 11, Voyager 2, and Uranus during 1985-1986. The latter two bodies are close to the ecliptic plane and hence their relationship is well represented by this diagram. However, Pioneer 11 is substantially north of the ecliptic plane, having a Z-coordinate of 5.753 AU on January 24, 1986 [Van Allen, 1984].

Desch and Kaiser [1984] and Hill and Dessler [1985] have considered, from different points of view, the prospects for detection of non-thermal radio emission with the planetary radio astronomy (PRA) instrument on Voyager 2 as it approaches Uranus. To my knowledge, no such emission has been identified as of the date of this writing.

6. THE MAGNETOSPHERE OF NEPTUNE

As previously mentioned, Pioneer 10 has established the flow of the solar wind out to and beyond the orbit of Neptune. By Figure 1, an empirically reasonable guess for the magnetic moment of Neptune if 6×10^{27} gauss cm^3 or 0.4 gauss a_N^3 and the corresponding stand-off distance is some 32 planetary radii.

Relevant comments are as follows:

(a) It seems probable that Neptune has a well developed magnetosphere.

(b) The rotational axis of Neptune is inclined at only 29° to the pole of its orbital plane. Hence, its magnetosphere may be expected to be more nearly "normal" than that of Uranus, perhaps most nearly resembling that of Saturn, except for the apparent absence of dense rings and close satellites.

(c) Of the two well-known satellites of Neptune—Triton and Nereid—only Triton is both close enough (14.6 planetary radii) and large enough (radius = 1750 km) to have a significant role in the planet's magnetosphere. But in contrast to Titan, Triton has, at most, a very tenuous atmosphere according to present evidence. (The tentatively identified satellite or partial ring, at ~ 3 planetary radii [Reitsema et al., 1982], may also qualify as an object of magnetospheric significance.)

(d) Very energetic particles from the cosmic ray albedo neutron source may be expected to dominate the inner magnetosphere as they do at Saturn.

(e) There is little prospect of substantial evidence on the magnetospheric properties of Neptune before the Voyager 2 encounter in August 1989.

7. CONCLUSION

Let me conclude by again recalling Frank McDonald's implicit admonition to be skeptical of any forecasts of magnetospheric properties, including those that I have just made.

ACKNOWLEDGMENT

This review was supported in part by Ames Research Center/NASA Contract NAS 2-11125 and by the U.S. Office of Naval Research Contract N00014-76-C-0016.

REFERENCES

Alfvén, H., 1950, *Cosmical Electrodynamics* (London: Oxford at the Clarendon Press).

Barnes, A., and Gazis, P., 1984, in *Uranus and Neptune*, NASA CP-2330, pp. 527-40.

Caldwell, J., Wagener, R., and Owen, T., 1983, *Nature*, **303**, 310-12.

Chapman, S., and Bartels, J., 1940, *Geomagnetism, Vol. I and II* (London: Oxford at the Clarendon Press).

Chang, A. F., and Lanzerotti, L. J., 1978, *J. Geophys. Res.*, **83**, 2597-2602.

Clarke, J. T., 1982, *Astrophys. J.*, **263**, L105-L109.

Dermott, S. F., 1984, in *Uranus and Neptune*, NASA CP-2330, pp. 377-404.

Desch, M. D., and Kaiser, M. L., 1984, *Nature*, **310**, 755-57.

Durrance, S. T., and Moos, H. W., 1982, *Nature*, **299**, 428-29.

Elliot, J. L., 1984, in *Uranus and Neptune*, NASA CP-2330, pp. 575-88.

Hill, T. W., and Dessler, A. J., 1985, *Science*, **227**, 1466-69.

McDonald, F. B., Schardt, A. W., and Trainor, J. H., 1980, *J. Geophys. Res.*, **85**, 5813-30.

Reitsema, H. J., Hubbard, W. B., Lebofsky, L. A., and Tholen, D. J., 1982, *Science*, **215**, 289-91.

Siscoe, G. L., 1971, *Planet. Space Sci.*, **19**, 483-90.

Siscoe, G. L., 1975, *Icarus*, **24**, 311-24.

Störmer, C., 1955, *The Polar Aurora* (London: Oxford at the Clarendon Press).

Van Allen, J. A., 1977, in *Highlights of Astronomy, Vol. 4* (International Astronomical Union).

Van Allen, J. A., 1984, "Geometrical Relationships of Pioneer 11 to Uranus and Voyager 2 in 1985-86," University of Iowa Research Report 84-14.

N87
24249
UNCLAS

2

AURORAL PARTICLES

David S. Evans
Space Environment Laboratory
National Oceanic and Atmospheric Administration
Boulder, Colorado 80303
and
Lockheed Palo Alto Research Laboratory
Palo Alto, California 94304

1. INTRODUCTION

While aurora have been a fashionable subject for study since the advent of the space age, a great deal of insight as to the origin of aurora had been gained in the years between the turn of the century and the Second World War, especially by Scandinavian physicists. It is illuminating to begin by reviewing their work as an illustration of the progress that can be made toward understanding a difficult subject through careful analysis and interpretation of observations. Although their methodology was based on the most advanced technology of the time, it would be considered entirely inadequate by today's standards. The background for this introduction may be found in books by Harang [1951], Störmer [1955], Chamberlain [1961], and Eather [1980].

Störmer's program of auroral photography, begun shortly after 1900, was directed toward defining the geometric shapes and locations of auroral forms. The best remembered of his experimental results concerned the distribution in altitudes of auroral features as derived from triangulation of photographic images taken of the same form against a star background from widely separated

points on the ground. As a general rule, the feature most easily identified as being common to the same form was the lower altitude border, although triangulations on other features, such as the high altitude boundary of extended rays, were also performed. The results showed that the lower altitude border of discrete auroral forms tended to have a distribution which maximized at about 105 km, with individual measurements extending down to 65 km and upwards to many hundreds of kilometers. During the same era, Vegard, also of Norway, made observations of the vertical extent of auroral luminosity and found these dimensions averaged some tens of kilometers but were hundreds of kilometers high in exceptional cases. Vegard also measured the vertical distribution of luminosity along an auroral form and showed that normally there was a maximum at an altitude some 10 km above the lower border.

Relatively little seems to have been done early in this century to quantify the horizontal dimensions of auroral forms. It was, however, clearly appreciated that the dimensions were usually very long in one direction (hundreds of kilometers east-west) and very thin in the other (tens of kilometers down to a small fraction of a kilometer for the structure in an individual auroral ray).

The period between 1890 and 1910 also saw the discovery of subatomic charged particles and radioactivity. Laboratory experiments had been conducted to study the properties of cathode rays (electrons), their penetrating power through materials, and to demonstrate how the rays interacted with a magnetic field. Similar experiments studied the nature of the particles emitted from radioactive materials, named alpha and beta rays, as well as the properties of fast ionized hydrogen atoms. Much work was also done with glow discharge tubes where fast subatomic particles interacted with gases resulting in the emission of light. Before 1900, Birkeland suggested that the aurora was produced by such subatomic particles transiting from the Sun to the Earth. Not only would this explain the emission of the light by processes analogous with those occurring in a gas discharge tube, but would also account for the observation that auroral rays aligned themselves along the geomagnetic field [a fact first noted by Wilcke in 1977] because these charged particles would be guided into the atmosphere along the nearly vertical magnetic field. The proposal

that the aurora was due to the impact of subatomic charged particles upon the atmosphere was generally accepted within a short time.

The altitude measurements of the aurora could be combined with the laboratory observations of the stopping power of subatomic particles passing through gases and with rudimentary models of atmospheric densities at high altitudes (based upon hydrostatic equilibrium arguments) to obtain estimates of the energies required for the incident particles to penetrate to the observed altitudes. Calculations done by Lenard, Vegard, and others before 1920 suggested that if the responsible particles were cathode rays, then the energies required to penetrate to 100 km would be on the order of 10 keV. If the particles were protons, the energies would be on the order of 200 keV. If alpha particles were hypothesized, the energies would need to approach 1000 keV. However, it was generally believed at this time that the responsible particles were cathode rays. There were two reasons for this. The first involved the height-luminosity profile of the aurora. If positive particles were responsible, then the luminosity profile of the aurora should exhibit a dramatic brightening with decreasing altitude and a very well-defined lower border, features that replicate the energy loss characteristics of positive particles passing through gases. These features were generally not present in the auroral luminosity profile which spoke in favor of cathode rays as being responsible. The second reason centered around the horizontal dimensions of auroral structures which were often very small. It was argued that these dimensions should in some respect be related to the gyro-radius of the responsible particles as they moved downward in the geomagnetic field. It was often impossible to reconcile the small horizontal dimension, which suggested that if massive positive particles were involved, they were relatively low velocity, with the altitude of the aurora which would require fairly energetic positive particles. If cathode rays were responsible, there would be no problems of this sort.

For a variety of reasons, speculation about the ultimate origin of the cathode rays centered around the Sun. First of all, evidence pointed toward an extraterrestrial source because auroras were almost exclusively phenomena occurring at high geomagnetic latitudes, and particles approaching the Earth from infinity would naturally be guided to high latitudes and excluded from low

latitudes by the geomagnetic field. Secondly, it had long been noted that the occurrence of auroras followed events on the Sun. The frequency of auroras duplicated the 11-year cycle in sunspot numbers and, equally important, unusually intense aurora and geomagnetic disturbances often followed solar flares by one or two days. Finally, correlations between the passage of sunspot groups past the central meridian of the Sun and aurora at the Earth showed that increases in auroral occurrence followed the sunspot passage by a day or so. By 1920, the picture emerged of the Sun ejecting streams of particles which transited to the Earth where they produced auroral displays. It was with this picture in mind that Störmer began his study of the orbits of charged particles moving from infinity into a dipole magnetic field. His objective was to demonstrate that particles originating from the Sun would impact the Earth's atmosphere at locations where the aurora was observed. His orbit tracing showed that particles (electrons or ions) of the energies believed necessary to penetrate to the proper depths in the atmosphere would impact at locations very close to the magnetic pole and not at the most frequent latitude for auroras which is considerably displaced from the pole. Stormer attempted to escape this problem by modifying the geomagnetic field through the addition of a toroidal current around the Earth (the ring current), but, even then, he could not obtain acceptable results. A second problem that existed with the picture had to do with the one-day delay between an event on the Sun and the subsequent aurora at the Earth. If this time were ascribed to the transit time for the particles responsible for the aurora, the particle velocities and energies were far too low to be identified with the supposed auroral particles; e.g., 10 eV electrons based upon transit time arguments against 10 000 eV electrons inferred from the altitudes of aurora and similar discrepancies, if the heavier positive ions were assumed.

By the start of the Second World War, it is fair to say that the following was thought to be "known" about the aurora. The aurora was caused by the impact of charged particles upon the atmosphere. These particles were probably electrons (cathode rays) and had energies on the order of 10 keV. The particles probably originated from the Sun and transited to the Earth in a time of about one day. It was appreciated, however, that this picture required the existence of an unknown near-earth process(es) which accelerated these particles to the energies required to produce the auroral form and that this same

process, or a second one, was responsible for establishing the geometric shapes and geographic location assumed by the aurora. Today, many of the justifications for current and future space plasma investigations are lineal descendants of these problems originally set down nearly 50 years ago.

2. PROGRESS FROM WORLD WAR II TO 1970

The first unambiguous evidence that energetic subatomic particles were participating in auroral displays was provided by Vegard and by Meinel in 1950. To the surprise of many, this evidence pointed toward energetic protons, rather than electrons, as the responsible particles. The observations were of line emissions from hydrogen atoms which had been Doppler-shifted to such an extent that the excited atoms must have been moving at thousands of km sec^{-1} (tens of keV energy) at the time of emission. Presumably, a proton entered the atmosphere with this magnitude energy, picked up an electron from within the atmosphere to become an excited neutral hydrogen atom, and had emitted the photon while still moving at high velocity. There was a period of time after these observations when it was thought by many that auroral light was produced primarily, if not exclusively, by proton bombardment of the atmosphere. This view did not last for long because the counter-arguments set down by the Scandinavian researchers before 1940 were far too compelling. Those original arguments, based upon the height-luminosity profiles and small scale dimensions often seen in auroral forms, had been bolstered by the fact that the intensity of the hydrogen emissions varied immensely compared to other auroral emission lines originating from normal atmospheric constituents such as atomic oxygen and molecular nitrogen. This could not be the case if proton bombardment were dominant.

The first direct measurements of the particles producing visible auroral displays were made by instruments on sounding rockets during the 1958 International Geophysical Year (IGY) program [Davis, Berg, and Meredith, 1960; McIlwain, 1961]. These rocket flights showed conclusively that the visible aurora was produced primarily by the precipitation into the atmosphere of electrons

having energies of the order of 10 keV. These rocket observations demonstrated that the energetic proton precipitation was responsible for only a small portion of the energy deposited into the atmosphere, and that the protons were incident over an area that extended well beyond that of the visible auroral forms.

The conclusions as to the nature of the particles causing the aurora that had been drawn by the Scandinavian physicists some 20 years earlier on the basis of indirect measurements were entirely vindicated by these rocket observations. McIlwain's work had particularly long lasting importance. The instrument on board the rocket used to measure the electron influxes was primitive by today's standards (the instrument was not capable of sensing electrons of energies less than 4 keV and only obtained rather crude energy flux versus electron energy distributions by means of a sweeping electromagnet) and the rocket performance was low, reaching only 120 km altitude. In spite of these limitations, McIlwain was able to combine the electron energy flux measurements with the altitude profile of the auroral luminosity obtained by a photometer on board the rocket to demonstrate that electrons of energies greater than 10 keV contributed less than 10% and electrons of energies less than 3 keV contributed less than 25% to the total particle energy flux incident upon the atmosphere. McIlwain characterized this electron energy distribution as being "near monoenergetic". He suggested further that this sort of energy distribution was not consistent with a "statistical type" acceleration mechanism, but that the "sharp high-energy cutoff" in the electron energy flux distribution was consistent with "acceleration processes involving electric fields".

Several years later and following the development of particle detectors better able to measure the fluxes of electrons over the energy range 100 eV to 10 keV, McIlwain's conclusion as to the near monoenergetic nature and origin of the auroral electron energy spectrum was fully accepted. Numerous measurements by rocket- and satellite-borne instruments have now shown that the type of electron energy spectrum first described by McIlwain is invariably observed above "discrete" auroral arcs. Figure 1, displaying the electron differential-directional number flux versus energy spectrum of the electrons observed over a bright auroral arc, is typical. Here, there is a peak in the distribution at about 10 keV with an extremely sharp decrease in intensity

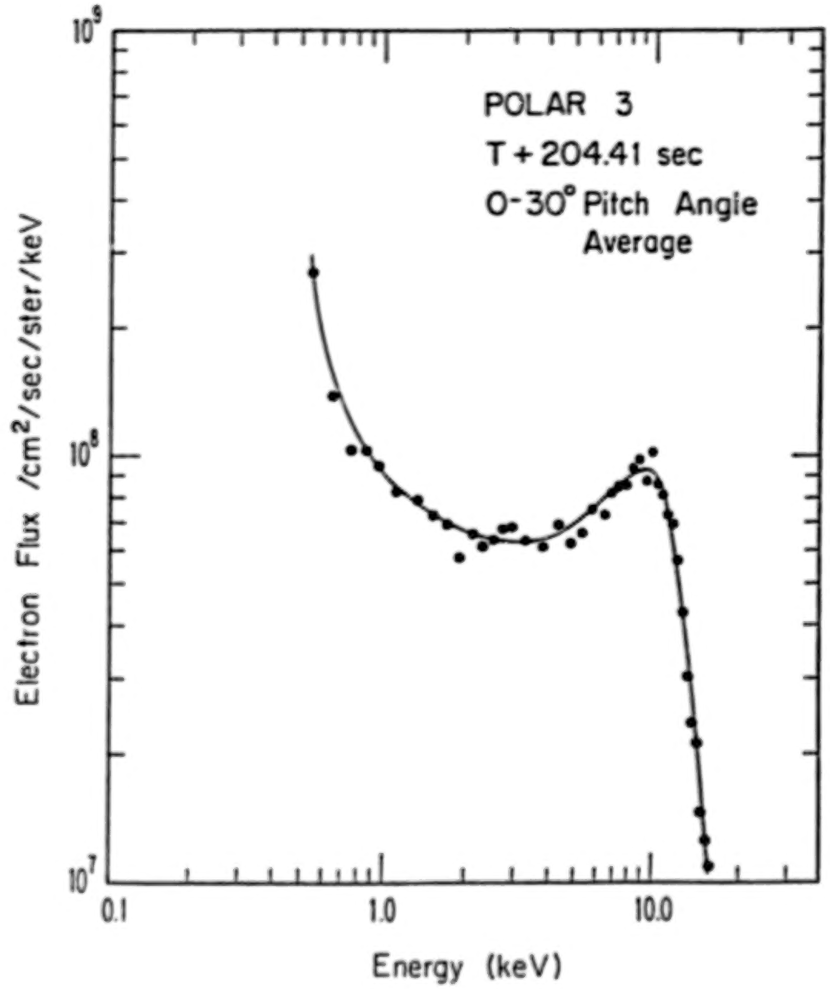


Figure 1. The electron number flux versus energy spectrum obtained above a discrete auroral arc. The spectral feature at near 10 keV is identified with the potential drop along the magnetic field line that accelerated the electrons. The lower energy electrons either originated from the atmosphere below the acceleration region or were accelerated electrons degraded in energy by some process. Positive ions played no role in the precipitation that produced this aurora.

at higher energies. While there are significant numbers of particles at lower energies, particularly below 1 keV, these particles contribute little to the incident energy flux. This form of energy spectrum may be fairly characterized as "near monoenergetic".

It should be pointed out that currently auroras are usually classified into two types, diffuse and discrete. As with most classification schemes, the separation is not totally unambiguous but does serve a useful purpose. Diffuse aurora tend to be widespread in spatial extent and uniform in intensity without well-defined boundaries to the light emission. It is this appearance which gives the name. Because of the absence of sharp boundaries, which would provide contrast against a dark sky background, this type of aurora is often difficult to recognize with the human eye. Moreover, the lack of easily identifiable features, such as rays or well-defined lower borders, make determinations of altitudes, locations, or height-luminosity profiles difficult to perform from the ground. For these reasons, relatively few rockets have been launched specifically to study such aurora, although satellite instrumentation has, on a regular basis, provided measurements of the particles producing such aurora.

Particle observations over diffuse aurora do show that energetic protons are participating in the bombardment, although it is rare that they carry more than 10% of the energy being deposited into the atmosphere, particularly when the total energy flux and auroral brightness are significant. The spectroscopic observations of Doppler-shifted hydrogen emissions made by Vegard, Meinel, and others were undoubtedly for cases of diffuse aurora. The energy spectra of both the electron precipitation responsible for the bulk of the energy deposition associated with the diffuse aurora and the lower intensity proton bombardment tend to resemble Maxwellian or "thermal-like" distributions without spectral features, such as the peak seen in the example in Figure 1. The mean energy of the electrons in the precipitation is ordinarily about 2 to 10 keV, while the mean energy of the protons is usually higher by a factor of 2 to 5. It is generally accepted that the electron and proton precipitation associated with the diffuse aurora originate from a reservoir of energized particles trapped in the geomagnetic field (the plasma sheet and outer radiation zones). These already energized particles are scattered in pitch angle by fluctuating electric fields and placed on trajectories leading into the atmosphere. The particle number densities and mean energies necessary on the part of the source plasma

population to supply the fluxes observed above diffuse aurora are in good agreement with the observed properties of the particle population in the plasma sheet. The contribution of protons to the total particle energy flux producing the diffuse aurora is also in quantitative agreement with this picture. For a source population composed of protons and electrons having equal densities and mean energies, and both species undergoing pitch angle scattering, the proton precipitation would account for about 2.5% of the total energy flux. If the proton temperatures were four times higher than the electron, the protons would contribute 5% under the same circumstances.

This explanation for the immediate origin of the particles producing the diffuse aurora does not address the question of the process whereby they acquired their energy. However, the Maxwellian-like nature of their energy spectrum suggests that the particles underwent collisional or randomizing processes as they were energized or afterwards. This would indicate an energization process which was either inherently statistical or, alternatively, a more ordered acceleration taking place in conjunction with statistical processes such as energy diffusion.

In contrast to the diffuse aurora, the discrete aurora has well-defined boundaries where the luminosity typically changes by a factor of 5 or more over a distance which is small compared to the overall dimensions of the form, a dimension which, in turn, is small compared to the dimensions of the diffuse aurora. The fact that discrete auroral forms exhibit such good contrast against a black sky allows easy identification of individual features for location and altitude determination. Most of the lower border altitude studies done by Störmer must have been of discrete forms. The excellent contrast of discrete auroras against a black sky combined with the fact that the amount of particle energy influx and auroral brightness is generally larger than for the diffuse aurora makes this type of aurora much easier to see and photograph. For these reasons, the bulk of ground-based studies have probably been performed on discrete aurora, and the majority of rockets have been launched over such aurora.

Measurements of the particles producing discrete aurora invariably show that precipitating electrons, of energies seldom exceeding 20 keV, are the dominant contributors to the energy deposition. The contribution due to proton

bombardment is virtually always less than 1% and often less than 0.1%. Moreover, the electron energy spectra essentially always show a peak or knee at some energy (Figure 1) or, on rare occasions, more than one energy. When displayed in terms of energy flux versus energy (Figure 1 displays number flux versus energy) the peak in the electron distribution is usually very dramatic. McIlwain's rocket flight, which encountered a near monoenergetic electron population, was over a discrete aurora. The nature of the electron energy spectrum together with the almost complete absence of protons contributing to the energy input has led most to conclude, as McIlwain did, that these electrons had acquired their energy by falling through an electrical potential difference. McIlwain did not speculate on the geometry of this potential difference or on the trajectory these electrons must have gone along between their arrival near the Earth (presumably from the Sun) and their deposition into the atmosphere. The search for answers to these questions have engendered considerable controversy in recent years.

3. THEORETICAL EXPLANATIONS FOR PARTICLE ENERGIZATION NEAR THE EARTH

If it is assumed that electrical potential differences exist within the magnetosphere surrounding the Earth, essentially three alternative geometries can be envisioned. The first is that the potential is distributed so that its gradient, the electric field, is everywhere perpendicular to the geomagnetic field. The second is a geometry in which some portion, if not all, of the available electrical potential is distributed parallel to the geomagnetic field. The third is a situation where the electric fields exist in a region where the magnetic field is zero, or nearly so.

Taylor and Hones [1965] explored the motion of charged particles in an electric and magnetic field geometry where the electric field was everywhere normal to the magnetic field. At that time, little was known of the nature of electric field that might surround the Earth and Taylor and Hones were compelled to develop a model based upon the electric fields that would be required to "drive" the currents known to flow in the ionosphere during times of magnetic

activity. The magnitude of these currents was estimated from the magnetic field perturbations that were observed from the ground and coupled with estimates of the electrical conductivity of the ionosphere to obtain the model of electric fields at ionospheric altitudes. These model electric fields were mapped upwards into the magnetosphere using a model magnetic field and assuming that the electric field always remained normal to the magnetic field. The creation of a model magnetic and electric field geometry for the entire magnetosphere was a major achievement for that time. Taylor and Hones then assumed the existence of a population of low energy particles at the boundary of their model magnetosphere (these particles having come from the Sun) and followed their trajectories in the model fields. In this particular geometry, the motion of individual low energy particles is a combination of a magnetically controlled drift, due to field line curvature and to gradients in the magnetic field, and an electrically controlled $\mathbf{E} \times \mathbf{B}$ drift. Given the gradients in the electric and magnetic fields and the scale size of the particle's gyroradius, the particle motion is adiabatic. The $\mathbf{E} \times \mathbf{B}$ drift alone is incapable of energizing particles because the drift path would be along an equipotential surface normal to $\mathbf{E}$. However, the superposition of the $\mathbf{E} \times \mathbf{B}$ drift and the magnetically controlled drift could carry the particles along a trajectory that has a component parallel to the electric field and result in the energization of the particle, effectively by moving through a potential difference. Particles of rather low solar wind energies can move through this geometry to a point where the electrostatic potential differs considerably from that at the entry point. The particle at this location will have its original energy plus that obtained by moving through the potential difference. If the latter exceeds the former by a significant amount, the particle energy spectrum at the final location will appear to be monoenergetic. If an individual particle is not precipitated into the atmosphere during its transit through the magnetosphere, its trajectory, being adiabatic, will return it back to the solar wind with its original energy. It should also be noted that the model leads to a separation between the trajectories followed by electrons and those followed by protons.

The model of Taylor and Hones explains, in a natural manner, both near monoenergetic particle spectra and the absence of energetic protons in the precipitation responsible for discrete auroras. However, the model does not easily account for the location and geometry of discrete auroral arcs. For a

particle source having a full range of incident pitch angles and energies at the magnetospheric boundary, the resultant trajectories do not form a line similar to an auroral arc but, rather, fill up much of the outer magnetosphere. The magnetosphere effectively acts as a crossed field particle analyzer with individual particles proceeding along trajectories to locations which are determined by their initial conditions (angle, velocity, mass, and charge). Taylor and Hones invoked a localized pitch angle scattering process, similar to that associated with the diffuse aurora, to precipitate the pre-energized particle population in the geometry appropriate to the discrete aurora. While this theory clearly describes one particle energization process which operates within the magnetosphere, it cannot easily account for these electrons producing discrete auroral arcs.

Speiser [1965, 1967] constructed a model in which an electric field was applied in a region of space, the magnetic tail, where the magnetic field was very small (a neutral sheet). Under these conditions, whether the electric field was perpendicular or parallel to the magnetic field was a moot point. Because in a near-zero magnetic field the dimensions of the particle's orbit would be large compared to the gradients in the electric and magnetic field, the motion of a particle would no longer be adiabatic. Using a tail-like magnetic field geometry and a dawn-to-dusk applied electric field, Speiser solved for the particle trajectories analytically. The results showed that a particle introduced into this geometry would undergo energization by moving parallel to the electric field while, at the same time, a north-south oscillatory motion between the tail lobes due to the particle's motion in the very weak neutral sheet magnetic field. Ultimately, the particle would either exit the system on the dawn (for electrons) or dusk (for positive ions) flanks, having been energized, or find itself in the tail lobes at a small pitch angle with respect to the magnetic field. Speiser showed that in the latter situation the particle, now energized, would follow a path along the magnetic field line toward the atmosphere. Speiser's model predicted near monoenergetic particle beams incident upon the atmosphere and that the electron and ion precipitation would be separated from one another. However, the model had difficulties in accounting for electrons of energies up to 10 keV without somewhat unrealistic assumptions about the magnetic field geometry. Essentially the electrons would be ejected from

the acceleration region very quickly and gain little energy from the electric field. Magnetic field configurations that would permit greater electron acceleration would result in those electrons entering the atmosphere at locations well poleward of where discrete aurora are usually observed. Finally, monoenergetic proton beams were also predicted but not observed even in proton-rich diffuse aurora.

Recently Lyons and Speiser [1982] expanded upon Speiser's original work and showed that if a plasma distribution having the number densities and temperatures of the plasma found in the "plasma mantle" were introduced into the neutral-sheet-electric-field acceleration geometry proposed by Speiser, the resultant ejected proton population calculated from the particle trajectories would have the intensities and energy distribution of those protons actually observed to be flowing on the outer edge of the plasma sheet. It is this population of positive ions that may be the major source of the plasma sheet population and, possibly, a direct source of auroral proton precipitation. The original Speiser model, perhaps in conjunction with additional particle energization by the Taylor and Hones adiabatic particle motion, may very well account for the energetic plasma population that forms the particle reservoir for the diffuse aurora precipitation where a significant admixture of protons is normally found. These same models, however, had difficulty in explaining the spatially structured, proton poor, and monoenergetic electron-rich character of the discrete aurora.

The third electric field acceleration geometry is one in which the electric field is directed parallel to the magnetic field. On the surface, this is a pleasing explanation. Assuming that the electric field is in the direction to accelerate electrons downward, electrons introduced across the high altitude boundary of the electric field will be energized and precipitated into the atmosphere in one direct process. The time taken for an individual electron to undergo the process is only seconds as opposed to a somewhat longer time for the original Speiser process and much longer for the Taylor and Hones adiabatic acceleration. In this picture, the geometry, small structure, and behavior of the discrete aurora are simply a manifestation of those magnetic field lines that possess

a parallel electric field, the total potential involved, and the time variations in that potential that may exist. The electron-rich nature of the discrete aurora and the monoenergetic spectrum are both direct consequences of the acceleration mechanism. By coupling the acceleration and precipitation process, the parallel electric field avoids the requirement invoked by Taylor and Hones for spatially structured pitch angle scattering processes to introduce corresponding spatial structure into the auroral precipitation. The particle acceleration by parallel electric fields may be located near the Earth instead of in the neutral sheet at great distances from the atmosphere as in the case of the Speiser model. This avoids the problem of accounting for small scale structure, introduced by the energization process, being preserved over long distances as the particles transit to the atmosphere. In spite of these seeming advantages, the idea that an electric field parallel to the magnetic field caused the energization and precipitation of those electrons responsible for discrete auroral arcs met with considerable resistance.

O'Brien [1970] summed up many of the arguments that a parallel electric field could not be the mechanism that energizes auroral particles. One point O'Brien stressed was that positive ions and electrons were observed to precipitate simultaneously. While this is usually the case for those particles producing the diffuse aurora (now interpreted as due to the loss of already energized particles from a reservoir, for example, the plasma sheet), the nearly monoenergetic electrons producing the discrete aurora seldom are accompanied by significant numbers of positive ions and do not seem subject to O'Brien's objection. A second, more telling, point made by O'Brien involved the presence of electrons in the precipitation which had energies lower than the magnitude of the accelerating potential difference that might be inferred from the location of the peak in the electron energy spectrum (e.g., those electrons of energies less than a few keV in Figure 1 where an accelerating voltage of 10 kV may be inferred). If the 10 keV electrons had fallen through a potential difference of that magnitude, then it seems that the lower energy electrons must have originated from within the region of parallel electric field and had acquired only a portion of the total available potential. However, as O'Brien pointed out, if this were the case, the large number fluxes of the low energy electrons

(see Figure 1) would require a powerful but unknown source of particles because the electrons locally available from within the region of parallel electric field would rapidly become exhausted.

Evans [1974] met this latter objection by pointing out that if there existed a parallel electric field which accelerated electrons of magnetospheric origin downward into the atmosphere, this same electric field would also reflect downward all secondary and backscattered electrons produced from the atmosphere by that precipitation. Effectively, the observed down-going electron population would be a combination of magnetospheric electrons energized by the electric field and electrons originating from the atmosphere that had been reflected back downward by that field. Numerical models were presented by Evans (Figure 2) which showed good agreement between an observed electron energy spectrum and that predicted using a backscatter-secondary model and a primary beam produced by accelerating a Maxwellian magnetospheric population through a fixed potential difference. Other comparisons between observation and model were not nearly so good. Generally in such instances, the predicted low energy electron fluxes were too low, particularly over energies between 20% and 80% of the accelerating potential (the spectrum in Figure 1 would likely be such an instance). A possible explanation for the excess of electrons over this energy range would be a process whereby the downgoing energized electrons produced beam-plasma instabilities in the ionosphere and the turbulent wave fields associated with this instability would diffuse electrons in energy, both promoting ionospheric electrons up in energy and degrading beam electrons down in energy [Evans, 1976]. In any case, the Earth's atmosphere and ionosphere represent a copious source of low energy electrons which would be confined below a parallel electric field, and so the existence of such electrons is not inconsistent with electron acceleration through a parallel electric field.

While O'Brien's objections, based upon observational considerations, to the existence of a parallel electric field which accelerated auroral electrons can be largely countered, there were also strong theoretical arguments against the very existence of electric fields parallel to a magnetic field, especially in the presence of a population of charged particles which were free to move under the influence of that electric field. The argument was that a static electric field

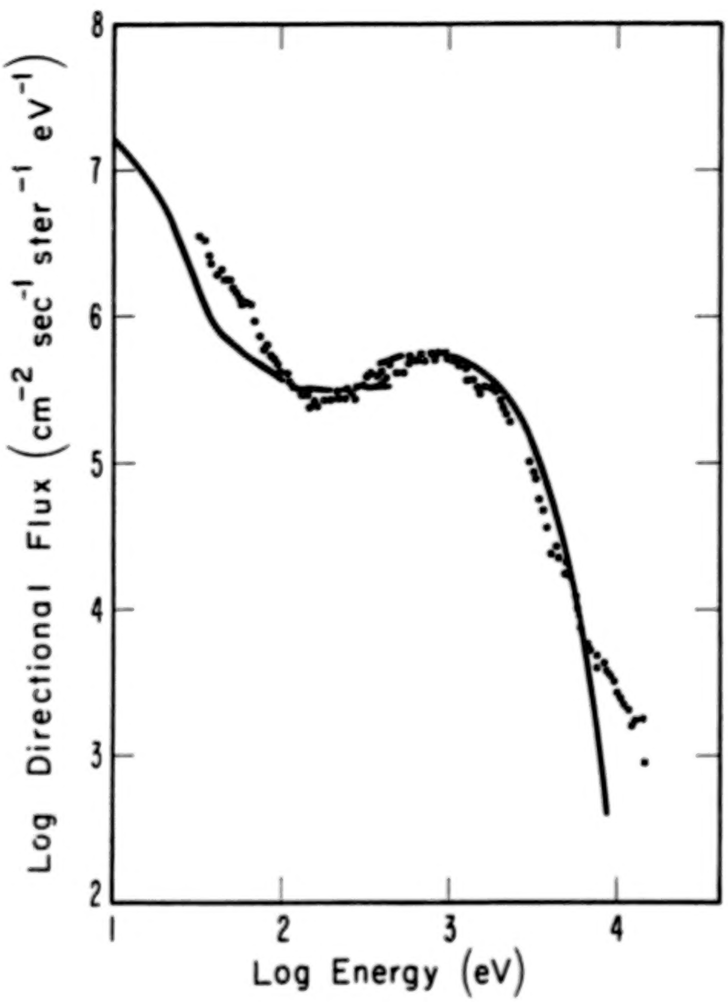


Figure 2. An example of a comparison between an observed auroral electron spectrum and one modeled by accelerating a Maxwellian distribution through a 500 volt field-aligned potential drop, computing the backscatter and secondary population created in the atmosphere by the precipitation, and reflecting that population from the potential barrier back downward. The good agreement overcame one of O'Brien's objections to the existence of parallel electric field acceleration.

parallel to a magnetic field also implied that a static charge distribution exists along the magnetic field line. However, the high mobility of electrons to move along the magnetic field (as opposed to across the field) under the influence of Coulomb forces would mean that such a charge distribution would be rapidly neutralized.

It was possible that an electrical current was flowing along the magnetic field line such that, as elementary charges moved out of a volume of space in a manner as to cancel the charge separation, new charges moved into that volume so as to maintain the charge distribution. In these circumstances, it was argued that the relationship between the current flowing along the magnetic field and the local electric field ought to be given by the Spitzer conductivity expression. Briefly, the Spitzer relationship between electric fields and currents had been derived in the following way. Consider a collection of ions and electrons, each with a number density, n, in the presence of an electric field. If there is a magnetic field also, it is assumed that the electric field is directed parallel to the magnetic field and the motion of charges in the direction normal to the magnetic field ignored. For simplicity, it is assumed that the ions are so massive in comparison to the electrons that their acceleration under the influence of electric forces is negligible and all current flow is due to the motion of electrons. The electrons are visualized as undergoing an acceleration due to the electric field but also a stopping or retarding force due to their occasional collisions with the massive slowly moving ions. This deceleration is expressed in this derivation as proportional to the electron's velocity $\vec{v}$ and a collision frequency, ν. Given this picture, the force balance equation may be written as

$$m\frac{d\vec{v}}{dt} = q\vec{E} - m\nu\vec{v} \qquad (1)$$

If a steady state is to be established, then the average acceleration $\frac{dv}{dt}$ can be set to zero and a steady drift velocity on the part of the electrons with respect to the ions would be given by:

$$\vec{v} = \frac{q}{m\nu}\vec{E} \qquad (2)$$

The electrical current, $\vec{J}$, produced by electrons of density, n, moving at this velocity is:

$$\vec{J} = nq\vec{v} \tag{3}$$

From equations (2) and (3) one obtains the Ohm's law relation between $\vec{J}$ and $\vec{E}$:

$$\begin{aligned} \vec{J} &= \frac{nq^2}{m\nu}\vec{E} \\ \vec{J} &= \sigma\vec{E} \end{aligned} \tag{4}$$

Equation 4 was believed to govern the relationship between electric fields and currents parallel to the magnetic field. The conductivity relating the two would be determined by the number densities of charged particles and the frequency of collisions by these particles as they moved. The conductivity would be very large at high altitude along the magnetic field line because moving charged particles would suffer collisions only infrequently in low density medium. The frequency of collisions would be greater in the denser medium at lower, ionospheric level, altitudes; but even here the electrical conductivity parallel to the magnetic field would be very much larger than conductivity perpendicular to the magnetic field—charged particles being confined to move along the magnetic field and inhibited from moving normal to that field. The very large electric conductivities parallel to the magnetic field suggested that no significant portion of any available electrical potential could appear parallel to the magnetic field, but virtually all such potential must appear transverse to the magnetic field. An alternative argument examined Equation 4 in the limit where the collision frequency approached zero. In this limit, both the conductivity and the field-aligned currents would grow unbounded if the parallel electric field remained non-zero. This result was regarded as unphysical and was thought to prove that any electric fields parallel to the magnetic field must remain very small or zero. A significant parallel electric field would exist only if the conductivity could be reduced by some process. Models which invoked scattering (collisions) of charged particles by interactions with turbulent electric fields (anomalous resistivity) were proposed to reduce this conductivity. The basic purpose of such a process was to inhibit the motion of

a charged particle along the magnetic field and so "support" an electric field without the field-aligned current growing unbounded. However, in order to account for the acceleration of auroral electrons (which was the purpose of proposing a parallel electric field in the first place), the concept of runaway electrons—electrons which did not interact with the turbulent fields but accelerated freely in the large scale parallel electric field—was invoked. The final picture appeared forced and unsatisfying.

There were several errors in the above analysis. One was that in order to set down the equation of motion (Equation 1), it was assumed that the motion of the charged particles in the electric field was collisionally dominated, i.e., that the energy gained by a particle from the electric field between collisions and the energy lost by the particle in a collision were both small compared to the thermal energy of the charged particle. The validity of the Ohm's law expression in Equation 4 was predicated on this assumption, and it was improper to allow the collision frequency to approach zero as this limit was completely out of the range of applicability of Equation 2. A second error was to assume Equation 3 was valid in the case where there were no collisions. In such a case, the current would not increase as the charged particles velocity increased but, rather, the number density, n, would decrease in exactly the same manner that the density of automobiles on a freeway decreases as their speed increases after escaping from a traffic jam. The current, $\vec{J}$, would remain the same and be governed not by the electric field but by the rate at which new charged particles were allowed to enter the system. Another very basic error in the formulation of these arguments was the presumption that the amount of current that flows at a given location was governed solely by the local electric field (or that the local electric field was determined by the current flow). In fact, the analysis does not explain why either the electric field or the currents have any particular values, or even why they should exist at a given location. The electrical currents flowing and the electric field existing at a given location are determined not only by properties at that location but also by the nature of the rest of the electrical circuit under consideration. For example, the charge exiting a given volume (the current) is determined in the steady state by the availability of charge in the adjacent volume and its ability to flow so as to maintain current continuity. The analysis presented above simply assumed that sufficient charges were available. If they were not, the

values of the current and electric fields (which are physically independent of one another) would simply change so that current continuity was once again established.

A far more acceptable approach to the entire problem of electric fields parallel to the geomagnetic field and any associated current systems has been put forward by Knight [1973], Lemaire and Scherer [1974], Fridman and Lemaire [1980], Chiu and Schultz [1978], Lyons [1980, 1981], and many others. This approach begins by examining the current that can flow along the magnetic field between the magnetosphere and ionosphere in terms of the ability of a particle population at one location to supply charge (a current) to another location. The charged particles responsible for a current flowing upwards from the ionosphere to the magnetosphere must either be positive ions from the ionosphere or electrons from the magnetosphere. A downward current from the magnetosphere to the ionosphere must be carried by ions originating from the magnetosphere together with electrons from the ionosphere. The ability of each of these plasma reservoirs to supply charged particles to the other depends first upon the number densities and temperatures of the available plasmas and second upon the ability of charged particles to transit from one location to the other without returning back along the same path.

As an illustrative example, consider the maximum charge flux (current) that can flow from one location to the other in the absence of any parallel electric fields. For the conventional upward current, the contribution of ions from the ionosphere (assuming a density of 1000 cm^{-3} and temperature of 1 eV at 1000 km) can be no more than about 0.7 μAmp m^{-2}, this maximum being the rate at which these ions can evaporate from the top of the ionosphere by virtue of their thermal motion. The contribution to this current from electrons originating from the magnetosphere can be no more than that flux given by electrons filling the loss cone and is about 1 μAmp m^{-2} (assuming a magnetospheric density of 1 cm^{-3} and temperature of 1000 eV), which gives a total maximum upward current that these two populations can supply to one another of about 2 μAmp m^{-2}. A similar analysis for the maximum downward current yields a value of about 30 μAmp m^{-2}. The order of magnitude difference between the maximum upward and downward charge fluxes (currents) that can flow between the magnetosphere and the ionosphere is due to the ionosphere's ability to supply an upward flux of electrons which

is much greater than the upward ion flux because of the low electron mass and high thermal speeds for a given temperature. Note that these estimates are for the maximum charge fluxes that can flow between the two regions in the absence of a parallel electric field, not the currents that actually do flow in any situation.

Knight [1973] and Lemaire and Scherer [1974] extended this sort of analysis to the situation where a magnetic field-aligned potential difference was assumed to exist at some altitude well above the ionosphere. Both the change in particle trajectories because of this assumed electrical potential difference and the new field-aligned current can be calculated. It is clear that there is no change to the maximum upward going charged particle fluxes because this value is determined by the "evaporation rate" of these particles from the ionosphere. The assumed parallel electric field may accelerate the ionospheric particles upward but cannot change the fluxes. The flux of particles from the magnetosphere to the ionosphere will be changed because the acceleration of particles downward will effectively widen the loss cone and more magnetospheric particles of the species determined by the direction of the parallel potential drop will reach the ionosphere. However, Knight and Lemaire and Scherer showed that this effect is not large. If the parallel potential drop were located just above the ionosphere, only magnetospheric particles already magnetically mirroring at low altitude would have their trajectories affected, and the charge flux would be little affected. If the potential were assumed to be at high altitude, well removed from the ionosphere, many magnetospheric particles would be affected, but the analysis showed there would be only a modest increase in the flux of particles actually reaching the ionosphere and, thus, the field-aligned current. Most of the magnetospheric particles, even with an acceleration downward, would still magnetically mirror above the ionosphere and would return back to the magnetosphere having been decelerated to their original energy by the field-aligned potential difference. The magnetic field-aligned currents that would flow between the magnetosphere and the ionosphere in the presence of a magnetic field-aligned electric field would not grow unbounded, as the arguments based upon conductivities in a collisionless plasma would have suggested, but would assume values which would be governed largely by the ability of particle populations outside the region of electric field to supply charge. This ability might be quite limited. The situation that would

exist is quite analogous to that of a thermionic diode where the currents that can flow between the cathode and anode are governed not only by the direction and magnitude of the electric field between those two surfaces but also by the ability of the cathode to make free charges available to flow.

Two other points should be noted about this picture. First, placing a parallel electric field in the ionosphere, where large numbers of charges are available to flow, will not enhance the current between the magnetosphere and ionosphere. This current is determined by the ability (and requirement) for charges to exercise a trajectory which carry them irreversibly from one region to another and this very low altitude electric field will not influence the particles in the magnetosphere nor increase the fluxes of ionospheric particles upwards which will still be given by the rate at which ionospheric particles can migrate into the electric field region ("evaporation rate"). Secondly, there is a clear asymmetry between the maximum downward current that can flow between the magnetosphere and ionosphere (magnetospheric ions transiting to the ionosphere and ionospheric electrons transiting to the magnetosphere) and the maximum upward current that could exist. In this respect, the system also mimics the characteristics of a thermionic diode, both in the unidirectional nature of the current flow and in the fact that placing an electric field along the wire leading to the cathode of the diode will not increase the current that can flow between cathode and anode.

Lyons [1980, 1981] made use of this analysis of the ability of currents to flow between the magnetosphere and ionosphere to develop a model which can account for the existence of an electrical potential difference along the magnetic field connecting these two regions, the acceleration of magnetospheric electrons, and the creation of discrete aurora. The model presumes that at high altitude in the magnetosphere there is an electric potential distribution imposed over a limited region of space and in a direction perpendicular to the magnetic field. This potential distribution represents a source of electromotive force (EMF) capable of providing a dissipative current which threads both the ionosphere and the source of EMF to the extent that currents can flow along the magnetic field connecting the two regions. The reasons for the existence of this potential distribution are not specified, although electric field

measurements, both in the ionosphere and in the magnetosphere, show such distributions must occur.

Lyons' model imposes the requirement for current continuity on a current system which flows through the source of EMF, along the field lines to the ionosphere, and closes by flowing horizontally in the ionosphere between those magnetic field lines carrying the upward and the downward currents. He demonstrates, given the nature of the ionospheric and magnetospheric charged particle reservoirs, that current continuity ordinarily cannot be established if one presumes that the magnetospheric potential distribution is mapped unaltered along the magnetic field lines between the magnetosphere and ionosphere (i.e., if there were no parallel potential difference). Essentially, if the magnetospheric electric field were mapped directly into the ionosphere, a large ionospheric current would result because of the immense number of charge carriers capable of moving horizontally at ionospheric altitudes. These large currents would be inconsistent with the field-aligned currents that could be carried by the available charged particles moving between the ionosphere and magnetosphere. An alternative distribution of the available magnetospheric potential around this current circuit involving field-aligned potential differences would be required. Indeed, Lyons' model shows that the major portion of the available potential must appear along the magnetic field line. This arises because even a large parallel potential difference will not produce a dramatic increase in field-aligned currents and so a major reduction in the potential difference across the ionosphere would be required to bring about current continuity.

It is natural in this picture for the parallel potential difference to appear on that leg of the current circuit which is required to carry an upward current. It is this leg that has the poorest current carrying capability. The sense of the parallel potential that would appear would be to accelerate electrons downward into the atmosphere which, of course, is exactly what is observed. It is satisfying that this picture explains why discrete auroral arcs are produced by downward accelerated electrons and seldom, if ever, by ions which had been accelerated downward by a parallel potential in the opposite sense. In the auroral current circuit, as in laboratory current circuits, potential differences arise in those regions where the current carrying capability is minimal.

Lyons' model says nothing about the distribution of the parallel potential differences (i.e., the parallel electric field), only the necessity for that potential difference and an estimate of its magnitude. The detailed potential distribution is a matter related to the microphysics which govern the motion of particles along the field line—including the effects of the parallel electric field. It is of interest to point out that magnetospheric electrons being accelerated downward toward the atmosphere and ionospheric ions accelerated upward produce, by virtue of the velocity changes on the part of these particles as they move, a space charge distribution along the magnetic field which is in the proper sense to be responsible for the potential distribution (viz. net negative space charge at high altitude and net positive space charge just above the ionosphere). This illustrates a point which seems little appreciated. In current carrying circuits, it is the charge carriers and the details of their motion that are responsible for distributing the electric fields that not only ensure current continuity but also locally govern the motion of the charge carriers themselves. The fact that the current carriers also play the role of the space charges responsible for the local electric fields which govern their own motion may appear paradoxical. However, the very existence of the dissipative current system requires a source of EMF, and the charge carriers are best viewed in terms of distributing this EMF around the circuit in this case rather than creating electric fields.

4. SUMMARY

The problems concerning the aurora posed prior to the war are now either solved in principle or have been restated in a more fundamental form. The Scandinavians thought the charged particles responsible for the aurora had come from the Sun. While strictly speaking this may not be entirely true (ionospheric ions accelerated upward by a parallel electric field may populate the magnetosphere and reappear as auroral particles; electron backscatter and secondaries from the atmosphere may undergo the same recycling), it is generally agreed that the energy required to create the aurora, and the various other dissipative processes associated with the aurora, comes from the Sun in the form of the kinetic energy of charged particles transiting the interplanetary medium. The pre-war hypothesis concerning the nature of the auroral particles and their energies has been fully confirmed, with the exception that

helium and oxygen ions (presumably of ionospheric origin) were identified as participating in the auroral particle precipitation in addition to the protons. The nature of the near-earth energization processes affecting auroral particles has been clarified. These processes involve electric fields, a fact which would not have come as a surprise to the pre-war physics community. Charged particle trajectories in various electric field geometries have been modeled. An electric field in a region of zero or very low magnetic field near the Earth is very effective in energizing particles and populating a reservoir with hot plasma but, perhaps, not so effective in setting these particles on trajectories which lead directly to the creation of aurora. An electric field everywhere perpendicular to the magnetic field also is effective in energizing plasma trapped in that magnetic field. One or the other or both of these near-earth electric field geometries seem quite capable of creating a population of energized plasma which, as the particles are precipitated into the atmosphere, would create the diffuse aurora.

It has also been shown that electric potential distributions imposed perpendicular to the magnetic field in the outer magnetosphere can lead to electric field distributions along a circuit path that threads through the ionosphere. The major portion of the available potential is along the magnetic field line linking these two regions. Moreover, the sense of this field-aligned potential difference develops preferentially to accelerate electrons from the magnetospheric reservoir of hot plasma downward into the atmosphere. This accounts for all the important characteristics of the discrete auroral display, particularly the monoenergetic nature of the electron energy spectrum and the relative lack of positive ion participation in the particle bombardment.

The physical problems have now moved from determining the nature and geometry of the electric fields, which accelerate charged particles near the Earth, to accounting for the existence of these electric fields as a natural consequence of the solar wind's interaction with the Earth. These explanations will undoubtedly center around such physical situations as the creation of charge separations, the exchange of particle kinetic energy and electromagnetic potential energy, and the character of electrical current systems in unbounded space.

It is my opinion that ultimately the reward in continuing the work in auroral and magnetospheric particle dynamics will be a deeper understanding of the

subtleties of classical electricity and magnetism as applied to situations not blessed with well-defined and invariant geometries. Many of the concepts currently held as valid may fail us in this problem, simply because those concepts were predicated on certain aspects of a physical situation, such as wires which predetermine current paths, that must be relaxed. We have already seen how the concept of conductivity misled us in the analysis of electric fields parallel to the magnetic field line in the presence of a collisionless plasma. The idea—that electrical charges moving around a circuit act not only as current carriers but also through their own motion as the agents responsible for distributing the electric field in the proper manner to ensure current continuity—has been clarified by consideration of auroral particle dynamics. Of course, this latter concept applies equally well in a laboratory circuit (as do all fundamental concepts in electricity and magnetism), although it is not emphasized because it seems unimportant to obtaining a solution to those problems. The unbounded space of the solar wind, magnetospheric, and ionospheric system is a problem in which all our familiar constraints must be relaxed. In this sense, it is a laboratory for the study of the interplay of mechanical and electrical processes in the purest of situations. As an understanding of this system is gained, it is inevitable that additional long believed concepts about the nature of electricity and magnetism in dynamical systems will need to be modified or discarded.

REFERENCES

Chamberlain, J. W., 1961, *Physics of the Aurora and Airglow* (New York: Academic Press).

Chiu, Y. T., and Schultz, M., 1978, *J. Geophys. Res.*, **83**, 629.

Davis, L. R., Berg, O. E., and Meredith, L. H., 1960, *Proc. COSPAR Space Science Symposium* (Amsterdam: North Holland Publishing Co.).

Eather, R. H., 1980, *The Majestic Lights* (Washington, DC: American Geophysical Union).

Evans, D. S., 1974, *J. Geophys. Res.*, **79**, 2853.

Evans, D. S., 1976, *Physics of Solar Planetary Environments*, ed. D. J. Williams (Washington, DC: American Geophysical Union).

Fridman, M., and Lemaire, J., 1980, *J. Geophys. Res.*, **85**, 664.

Harang, L., 1951, *The Aurorae* (New York: John Wiley & Sons).

Knight, L., 1973, *Planet. Space Sci.*, **21**, 741.

Lemaire, J., and Scherer, M., 1974, *Planet. Space Sci.*, **22**, 1485.

Lyons, L. R., 1980, *J. Geophys. Res.*, **85**, 17.

Lyons, L. R., 1981, *J. Geophys. Res.*, **86**, 1.

Lyons, L. R., and Speiser, T. W., 1982, *J. Geophys. Res.*, **87**, 2276.

McIlwain, C. E., 1961, *J. Geophys. Res.*, **66**, 2727.

O'Brien, B. J., 1970, *Planet. Space Sci.*, **20**, 1821.

Speiser, T. W., 1965, *J. Geophys. Res.*, **70**, 4219.

Speiser, T. W., 1967, *J. Geophys. Res.*, **72**, 3919.

Störmer, C., 1955, *The Polar Aurora* (London: Oxford University Press).

Taylor, H. E., and Hones, E. W., 1965, *J. Geophys. Res.*, **70**, 3605.

N87
24250
UNCLAS

N87-24250

3

PARTICLE ACCELERATION IN SOLAR FLARES

R. Ramaty
Laboratory for High Energy Astrophysics
Goddard Space Flight Center
Greenbelt, Maryland 20771

M. A. Forman
Department of Earth and Space Sciences
State University of New York
Stony Brook, New York 11794

ABSTRACT

The most direct signatures of particle acceleration in flares are energetic particles detected in interplanetary space and in the Earth's atmosphere, and gamma rays, neutrons, hard X-rays, and radio emissions produced by the energetic particles in the solar atmosphere. We review the stochastic and shock acceleration theories in flares, and we discuss the implications of observations on particle energy spectra, particle confinement and escape, multiple acceleration phases, particle anistropies, and solar atmospheric abundances.

1. INTRODUCTION

Acceleration of energetic particles is a widespread phenomenon in nature, one that occurs at a variety of sites ranging from the Earth's magnetosphere to distant objects such as supernovae, active galaxies, and quasars. There are, in fact, many explosive phenomena in astrophysics, solar flares among them, in which energetic particles are routinely produced and often contain a large fraction of all the available energy.

It is widely believed [e.g., Syrovatskii, 1981] that solar flares draw their energy from the annihilation of magnetic fields. The rapid energy deposition following the annihilation should be an important source of turbulence and shocks. As proposed by Fermi [1949], charged particles can be accelerated to high energies in repeated reflections from magnetized clouds, or, in the more recent view, from hydromagnetic turbulence and shocks. This mechanism must be important in solar flares where shocks are known to exist and turbulence is expected to be produced by both shocks and other mechanisms. In addition, particles may also be accelerated in the strong electric fields which should accompany the annihilation of magnetic fields.

Indeed, there is ample evidence for particle acceleration in flares. Energetic particles accelerated at the Sun are directly detected in interplanetary space and the neutral radiations (radio emissions, hard X-rays, gamma rays, and neutrons) produced by the accelerated particles in the solar atmosphere are observed with detectors on the ground and in space. These observations provide a variety of information on the properties of the particles, including their energy spectrum, total numbers at the Sun and in interplanetary space, and angular distribution at the Sun. In addition, the measured composition of solar flare particles in interplanetary space [e.g., Breneman and Stone, 1985] and flare gamma ray line spectroscopy [Murphy et al., 1985] provide new information on the chemical composition of the solar atmosphere.

In a previous paper [Forman, Ramaty, and Zweibel, 1986, hereafter FRZ] we have reviewed much of the data, as well as the stochastic, shock, and electric field theories of particle acceleration in solar flares. Several other recent reviews are also available, e.g., Goldman and Smith [1986] on radio emissions, Dennis [1985] on hard X-ray bursts, Hudson [1985] on ions in solar flares, and Lin [1985] on solar electrons in the interplanetary medium. Here we discuss two of the leading acceleration mechanisms, the stochastic Fermi process and diffusive shock acceleration. These mechanisms not only are expected to be important for particle acceleration in flares but also appear to be capable of accounting for many observed solar flare phenomena related to the acceleration of protons and relativistic electrons [Ellison and Ramaty, 1985; Murphy, Dermer and Ramaty, 1987]. We also discuss recent observations of charged particles, gamma rays, and neutrons, with particular focus on the simultaneous observations of flares in all three of these channels.

The present paper is the expanded version of the talk given by one of us (MAF) at Frank McDonald's sixtieth birthday symposium. Frank, through his extensive and pioneering observations of accelerated particles of planetary, solar, interplanetary, and galactic origins, has made fundamental contributions to the understanding of astrophysical particle acceleration. Frank's most recent contribution to solar flare studies, the observation of accelerated particles from two neutron producing flares [McDonald and Van Hollebeke, 1985], is fundamental to the considerations developed and discussed in the present paper.

2. ACCELERATION MECHANISMS

A. Stochastic Acceleration

Processes in turbulent plasmas which cause particles to change their energy in a random way with many increases and decreases in energy lead to stochastic acceleration. In the original stochastic Fermi mechanism [Fermi, 1949], the process was idealized as reflection from randomly moving magnetized clouds. Stochastic acceleration can also result from resonant pitch-angle scattering from Alfvén waves with wavelengths of the order of the particle gyroradius. To accelerate particles these waves must propagate both parallel and antiparallel to the average magnetic field [Skilling, 1975]. Other modes of stochastic acceleration, called magnetic pumping and transit-time damping, occur through interaction with magnetosonic waves whose wavelengths are much longer than the particle gyroradius [Kulsrud and Ferrari, 1971; Melrose, 1980; and Achterberg, 1981]. These modes require additional pitch-angle scattering to keep the particles isotropic. Langmuir (plasma) waves or other electrostatic waves with phase velocities of the order of the particle speed will also accelerate particles stochastically [Melrose, 1980].

When the random energy increments are small compared to the particle energy, stochastic acceleration can be described as diffusion in momentum space characterized by a momentum diffusion coefficient D_{pp}. Expressions for D_{pp} for the various processes mentioned above were summarized in FRZ. Stochastic-acceleration spectra can be obtained from these D_{pp}'s by solving a transport equation [see FRZ and Droge and Schlickeiser, 1986] which takes

into account the injection of the particles, diffusion in momentum space, nondiffusive energy changes (e.g., ionization losses, and nuclear collisions), and escape from the acceleration region. Several of the published stochastic-acceleration spectra are given in FRZ.

Here we discuss the simple model of acceleration by hard sphere scattering which has been applied extensively to both solar flare particle observations in interplanetary space [McGuire and von Rosenvinge, 1984] and gamma ray and neutron production models at the Sun [Murphy and Ramaty, 1984; Murphy, Dermer, and Ramaty, 1987]. In this treatment [Ramaty, 1979; FRZ] the scattering mean free path is assumed to be independent of particle energy and species, the acceleration region is characterized by a constant escape time and all additional losses are ignored. With a steady source of q particles cm^{-3} s^{-1} at an injection momentum p_0, the steady state particle density in phase space as a function of particle momentum p is given by:

$$f = 6\,q / (4\,\pi\,p_0^2\,mc\,\alpha)\,(p_0/p)\,\cdot$$

$$\cdot\, I_2(2(3p_0/(mc\,\alpha T))^{1/2})\,K_2(2(3p/(mc\,\alpha T))^{1/2}), \tag{1}$$

if both p and p_0 are nonrelativistic, and by:

$$f = q / (4\,\pi\,p_0^3\,\alpha\,(1+4/(3\,\alpha T))^{1/2})\,.$$

$$(p/p_0)^{-3/2-\sqrt{(9/4+3/\alpha T)}}\ , \tag{2}$$

when both p and p_0 are ultrarelativistic. Here m is particle mass, p particle momentum, T the escape time from the acceleration region, and $\alpha = (\delta V)^2/(\lambda\,c)$ where $(\delta V)^2$ is the mean square velocity of the scatterers and λ the diffusion mean free path in momentum space. The combination of parameters αT characterizes the shape of the spectrum such that a larger value of αT corresponds to a harder spectrum. The phase space density f is related to the differential particle density, N(E), or the differential particle flux, J(E), $N(E) = J(E)/v = A\,p^2\,f/v$, where E is energy per nucleon, A is particle mass number, and v particle velocity. In both Equations (1) and (2) p must be greater

than p_0. For $p < p_0$, the arguments of I_2 and K_2 in Equation (1) are interchanged, and the minus sign in front of the square root in the exponent in Equation (2) is changed to a plus sign.

The nonrelativistic expression, Equation (1), can be used to fit solar flare accelerated ion spectra up to about 100 MeV/nucleon. Such fits are discussed in Section 3. The ultrarelativistic expression, Equation (2), could be used to fit electron spectra above about 1 MeV. But if αT is rigidity independent, the values of αT that fit the proton spectra produce ultrarelativistic electron spectra which are much steeper than those observed [e.g., Ramaty, 1979]. Clearly, a complete stochastic acceleration model should take into account the rigidity dependence of αT, but no attempts have yet been made (see FRZ) to compare the theory with data in detail.

A very important question in all particle acceleration theories, including stochastic acceleration, is that of injection. We first note that the basic concept of stochastic acceleration assumes that the energy changes are small compared with the particle energy and therefore the particle velocity must be much greater than δV. Furthermore, for resonant scattering, ions must have $v > V_A$ to scatter from Alfvén waves and electrons must have $v > 43V_A$ to scatter from whistlers [Melrose, 1974]. An additional injection condition is set by the requirement that the systematic acceleration rate due to diffusion in momentum space be larger than the ionization and Coulomb energy loss rates of the particles. A detailed comparison of these rates was given in FRZ.

The acceleration time in stochastic acceleration can be studied from the time dependent solutions of the transport equation. The relevant parameters are δV and λ. If we set $\delta V = V_A$, then a typical value is $\delta V = 1000$ km/sec, corresponding to a magnetic field of 100 gauss and an electron density of 5×10^{10} cm^{-3}. The diffusion mean free path should be at least as large as the particles gyroradius. Taking λ to be 100 times the gyroradius of a 20 MeV proton, we obtain $\lambda = 7 \times 10^5$ cm. Then the acceleration parameter $\alpha \cong 0.5$. For this α, the time dependent solutions of the transport equation obtained [Ramaty, 1979; FRZ] for impulsive injection and perfect trapping ($T \rightarrow \infty$) imply that a significant number of protons are accelerated to tens of MeV in less than 1 sec. In general, a time of the order $1/\alpha$ is required to accelerate particles to relativistic energies.

B. Shock Acceleration

Solar flare shocks propagate upward through the solar corona at speeds of about 500 to 2000 km/sec, as indicated by Type II radio bursts [e.g., Goldman and Smith, 1986] and laterally through the chromosphere where they are seen as Moreton waves [Uchida, 1974]. The occurrence of solar energetic particles in space is strongly correlated with flares having Type II bursts [Svestka and Fritzova, 1974]. Flare shocks, corotating shocks, and planetary bow shocks are observed to accelerate particles in interplanetary space (see references in FRZ). A flare shock can transport particles in an energy independent manner through the corona until they escape onto open field lines. Shock acceleration has been reviewed by Toptygin [1980] and Axford [1982] and applied to solar flares by Achterberg and Norman [1980], Decker, Pesses, and Armstrong [1981], Ellison and Ramaty [1985], and Lee and Ryan [1986].

There are basically two types of shock acceleration: scatter-free, in which particles gain energy by reflection in a single shock encounter [Sonnerup, 1973; Armstrong, Pesses, and Decker, 1985; references in FRZ] and diffusive, in which particles gain energy by repeated scattering between the converging plasmas on either side of the shock [e.g., Axford, Leer, and Skadron, 1977; Axford, 1982 and; Forman and Webb, 1985]. The scatter-free mechanism can enhance the particle energy by about an order of magnitude if the shock is nearly perpendicular (i.e., the magnetic field is nearly perpendicular to the shock normal), but in that case only particles with speeds which are already much greater than the shock speed can be reflected. Further acceleration, however, requires multiple reflections. These are possible if there is particle scattering in the fluid flow or if the particles are trapped between converging shocks in a flare loop [Wentzel, 1965; Bai et al., 1983].

Acceleration by diffusive scattering across the shock is a first order Fermi process, in the sense that every shock crossing results in an energy gain. It is in principle more efficient than stochastic acceleration because it derives energy directly from the compression of the flow at the shock. For this mechanism to be effective, there must be adequate particle scattering both upstream and downstream of the shock. The passage of the shock is expected to generate turbulence downstream which will scatter the particles. Scattering upstream, however, is more problematic [Holman, Ionson, and Scott,

1979]. Observations [Tsurutani and Rodriquez, 1981] of interplantetary shocks and planetary bow shocks show that when they are nearly parallel there is a very turbulent foreshock region capable of scattering particles. Such a region could be produced by the accelerated particles themselves [Achterberg and Norman, 1980; Lee, 1982].

In the simplest example of diffusive shock acceleration at a planar shock where the only losses are due to convection of the particles away from the shock downstream, the energetic particle density in phase space is given by a power law in momentum, $f \sim p^{-\sigma}$, where $\sigma = 3V/\Delta V$, V is the shock speed and ΔV the discontinuity in the plasma speed at the shock. In terms of the shock compression ratio, r, $\sigma = 3r/(r-1)$. For a strong shock in a nonrelativistic fluid $r = 4$ and hence $f \sim p^{-4}$. For weaker shocks, $4 > r > 1$ and the power law is steeper. The momentum power law $p^{-\sigma}$ implies [Ellison and Ramaty, 1985] that differential flux as a function of energy per nucleon, J(E), has an energy dependent spectral index which approaches $s = (r+2)/(r-1)$ in the ultrarelativistic regime and s/2 in the nonrelativistic regime.

A variety of effects truncate the power law behavior of shock-accelerated spectra at high energies. The effects that have been considered specifically are adiabatic deceleration [Lee and Ryan, 1986], shock lifetimes comparable to particle acceleration times [Forman, 1981] and shock sizes comparable to particle diffusion lengths [e.g., Ellison, 1984]. These effects all produce spectra that turn over at high energies when the diffusion coefficient increases with momentum. The derivation of the exact forms of these turnovers generally requires numerical treatments. In many cases, however, they can be approximated by exponentials [e.g., Ellison and Ramaty, 1985],

$$f_i(p) \sim p^{-\sigma} \exp(-E/E_{0i}), \tag{3}$$

where f_i is the phase space density of particles of species i, and E_{0i} is the turnover kinetic energy for species i. In the case of particle escape at an escape boundary at a given distance from the shock, E_{0i} depends on the distance to this boundary and the diffusion coefficent. If the diffusion mean free path

is proportional to particle gyroradius, the turnover energies of the various species are related by:

$$v(E_{0i})\, R(E_{0i}) = v(E_{0p})\, R(E_{0p}), \tag{4}$$

where R is particle rigidity, E_{0p} is the proton turnover kinetic energy, and E_{0i} is energy per nucleon for nuclei and energy for electrons. Equation (3) can be used to fit both ion and electron spectra. We present such a fit in the next section.

As for stochastic acceleration, and for shock acceleration as well, the question of injection is very important. Ionization and Coulomb energy losses to the ambient medium have the same role in determining injection conditions in shock acceleration as they do in stochastic acceleration. In addition, for diffusive shock acceleration, particles downstream must have sufficient velocity to overtake the shock. This is at least $(V - \Delta V)$ directed towards the shock, and increases as $(\cos \psi)^{-2}$, where ψ is the angle between the downstream field and the shock normal. The velocity $V - \Delta V$ is at least as great as V_A or ΔV, and with the additional $(\cos \psi)^{-2}$ factor, the threshold for shock acceleration could be higher than for stochastic acceleration.

Another injection condition is set by the finite width of the shock which could depend on many parameters including the pressure of the accelerated particles. When this pressure is taken into account [Axford, Leer, and Skadron, 1977; references in FRZ] all or part of the velocity change ΔV is smoothed out over a length scale $\sim \bar{k}/V$, where $\bar{k}$ is the diffusion coefficient of the energetic particles averaged over their energy spectrum. This smoothing could affect the composition of the accelerated particles [Eichler, 1979]. Drury, Axford, and Summers [1982] show analytically that when k is independent of energy, the smoothing causes the dominant accelerated species (i.e., protons) to have a steeper spectrum than in the case of an infinitely thin shock. Minor species which are only partially stripped could have larger diffusion coefficients than the protons if the diffusion mean free path is rigidity dependent, and therefore for them $k > \bar{k}$. Drury, Axford, and Summers [1982] show that the spectrum of such minor species is flatter than for protons and approaches that of an infinitely thin shock.

The mean rate of energy gain in shock acceleration is (see FRZ):

$$dE/dt = 3/4\ (V\ \Delta V)\ (p/\lambda). \tag{5}$$

Ellison and Ramaty [1985] have shown that if λ is proportional to gyroradius, the acceleration time for both protons and electrons is:

$$t_a \cong 2\times 10^{-4}\ f\ (E - E_{inj})/((B/100\ \text{gauss})(V\ \Delta V/10^6\ \text{km}^2/\text{sec}^2)), \tag{6}$$

where t_a is in sec, E and E_{inj} are the final and injection energies in MeV, and f is the ratio of λ to the gyroradius. It is expected that f is much greater than unity. For $f = 100$, $B = 100$ gauss and $V = \Delta V = 10^3$ km/sec, the acceleration time to 20 MeV is ~ 0.4 sec. We note that the acceleration time is the same for protons and electrons, a result which follows from the assumed linear dependence of the mean free path on gyroradius.

3. OBSERVATIONS AND THEIR INTERPRETATIONS

We discuss the following: (1) energy spectra, (2) the trapping and escape of the particles and the related problem of multiple acceleration phases, (3) angular distributions at the Sun, and (4) abundance determinations from energetic particle and gamma ray observations.

A. Energy Spectra

The energy spectrum of accelerated particles is perhaps the most important diagnostic of acceleration mechanisms. Information on energy spectra is obtained from both the charged particles observed in interplanetary space, and the gamma ray and neutron data. The interplanetary observations provide information on the energy spectrum of the particles which escape from the Sun.

As indicated by upper limits on ^{2}H, ^{3}H, Li, Be, and B [McGuire, von Rosenvinge, and McDonald, 1979], the amount of matter traversed by the escaping particles (< 0.1 g/cm^{-3}) is generally insufficient for the production of the observed gamma rays [Ramaty, 1986]. Gamma rays and neutrons are produced by the particles which slow down and thermalize in the solar atmosphere.

Observations of these neutral emissions, therefore, provide information on the particles which remain trapped at the Sun.

The energy spectra obtained from charged particle observations have been reviewed by McGuire and von Rosenvinge [1984]. In determining spectra from such observations the effects of coronal and interplanetary transport must be minimized. It was shown [Van Hollebeke, MaSung, and McDonald, 1975] that this can be achieved by considering only particle events from flares that are well connected magnetically to the observing spacecraft and by constructing the energy spectra at times of maximum intensity at each energy. Using this technique, McGuire, von Rosenvinge, and McDonald [1981] and McGuire and von Rosenvinge [1984] showed that over the broad energy range (from about 1 to 400 MeV) proton spectra can be fit with the Bessel function spectra expected from stochastic acceleration [Equation (1)] with values of αT between 0.015 and 0.035. But as emphasized in FRZ, energy spectra obtained from shock acceleration could also fit these data.

Proton spectra have been determined from gamma ray and neutron observations using three different techniques [Murphy and Ramaty, 1984; Murphy, Dermer, and Ramaty, 1987]. The first one considers the ratio of the fluence in excess of a power law in the 4-7 MeV range to the fluence in the 2.223 MeV line. This excess is due to nuclear line emissions. The second technique uses observed neutron spectra which depend on the spectrum of the protons which produce the neutrons; and the third is based on the ratio of the 4-7 nuclear excess to the pion decay emission at photon energies > 10 MeV. The first technique is sensitive to the spectrum in the 10 - 100 MeV range, the second probes the spectrum in the 50 MeV - several GeV range and the third is sensitive to the ratio of the number of particles in the 10 - 100 MeV range to the number above the pion production threshold (few hundreds MeV).

When analyzed in terms of the Bessel function of Equation (1), the 4-7 MeV to the 2.223 MeV line ratios for 15 flares yielded values of αT in the range from 0.014 to 0.038 [Murphy and Ramaty, 1984; Hua and Lingenfelter, 1987a]. This range is essentially the same as that obtained from the interplanetary charged particle data. However, the implications of this overlap are not clear because most of the particle events correspond to a different set of flares than the one used to derive the gamma ray results.

There are, in fact, only three flares for which energy spectra have been determined simultaneously from gamma ray and charged particle data: June 7, 1980; June 21, 1980; and June 3, 1982. For the June 7 flare, the 4-7 MeV to 2.223 MeV ratio yields $\alpha T = 0.021 \pm 0.003$ [Murphy and Ramaty, 1984; Hua and Lingenfelter, 1987a] and the charged particle data yield $\alpha T = 0.013$ [McGuire and von Rosenvinge, 1985]. For the June 21 flare, the 4-7 MeV to 2.223 MeV ratio yields $\alpha T = 0.022 \pm 0.007$ [Hua and Lingenfelter, 1987a] and the charged particle data yield $\alpha T \cong 0.025$ [Murphy, Dermer, and Ramaty, 1987] using the data of McDonald and Van Hollebeke [1985] shown in Figure 1. For the June 21 flare αT was also determined [Murphy and Ramaty, 1984] from neutron observations [Chupp et al., 1982] yielding $\alpha T \cong 0.025$. We see that for both the June 7 and June 21 flares the spectrum of the particles which escaped from the Sun could have been similar to the spectrum of the particles which slowed down and thermalized in solar atmosphere. It is possible that for these flares a common acceleration mechanism operating at a common acceleration site is responsible for both particle populations.

The interplanetary proton and electron spectra from the June 3, 1982 flare [McDonald and Van Hollebeke, 1985] are shown in Figure 2 [from Murphy, Dermer, and Ramaty, 1987]. As can be seen, a power law fits the proton spectrum better than does a Bessel function. This power law is given by Equation (3) with $s = 2.4$ ($\sigma = 4.4$) and E_{0p} >>300 MeV. The gamma ray data from this flare, however, cannot be reproduced with a single, time independent, proton spectrum [Murphy, Dermer, and Ramaty, 1987]. Early in the flare (for about 100 sec) the proton spectrum can be fit with a Bessel function with $\alpha T \cong 0.04$, but later in the event the observed ratio of the 4-7 MeV nuclear excess [Prince et al., 1983; Chupp et al., 1987] to the emission from the decay of pions [Forrest et al., 1985, 1987] implies that the spectrum must harden. At these later times the spectrum is consistent with the proton spectrum observed in interplanetary space ($s = 2.4$ and $E_{0p} > 300$ MeV).

The inferred hardening of the proton spectrum suggests that not all the particles accelerated in the June 3 flare had a common origin. Murphy, Dermer, and Ramaty [1987] show that a wide variety of data from the June 3 flare (pion decay emission, nuclear deexcitation lines, the 2.223 MeV and 0.511

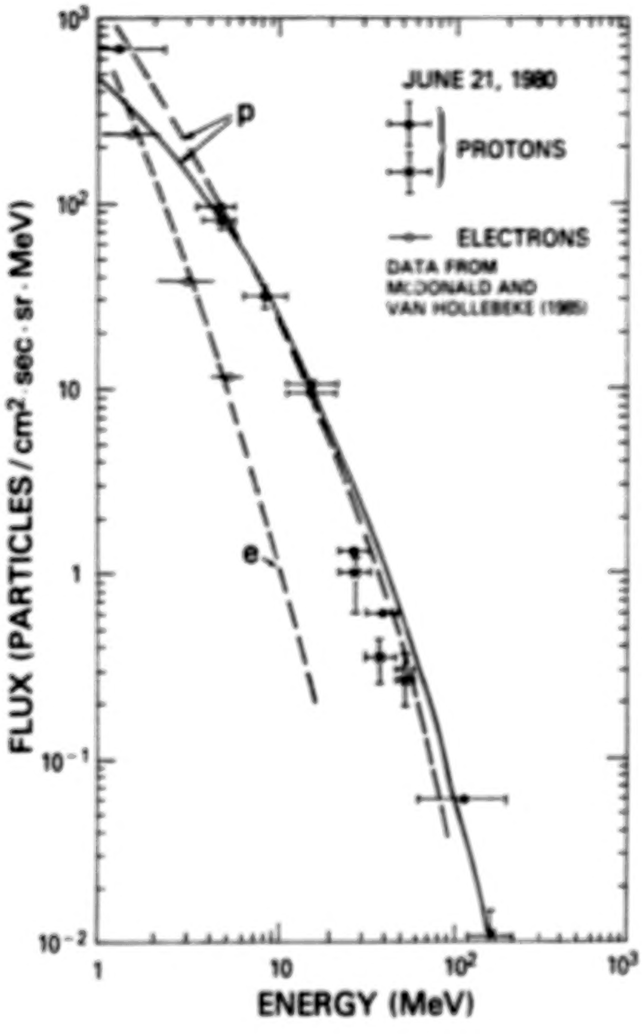


Figure 1. Interplanetary proton and electron observations [McDonald and Van Hollebeke, 1985] of the June 21, 1980 flare. The solid curve corresponds to a stochastic-acceleration spectrum with $\alpha T = 0.025$, and the dashed curves correspond to shock-acceleration spectra with $s = 3.3$ and $E_{0p} = 30$ MeV for the protons and $E_{0e} = 59$ MeV for the electrons.

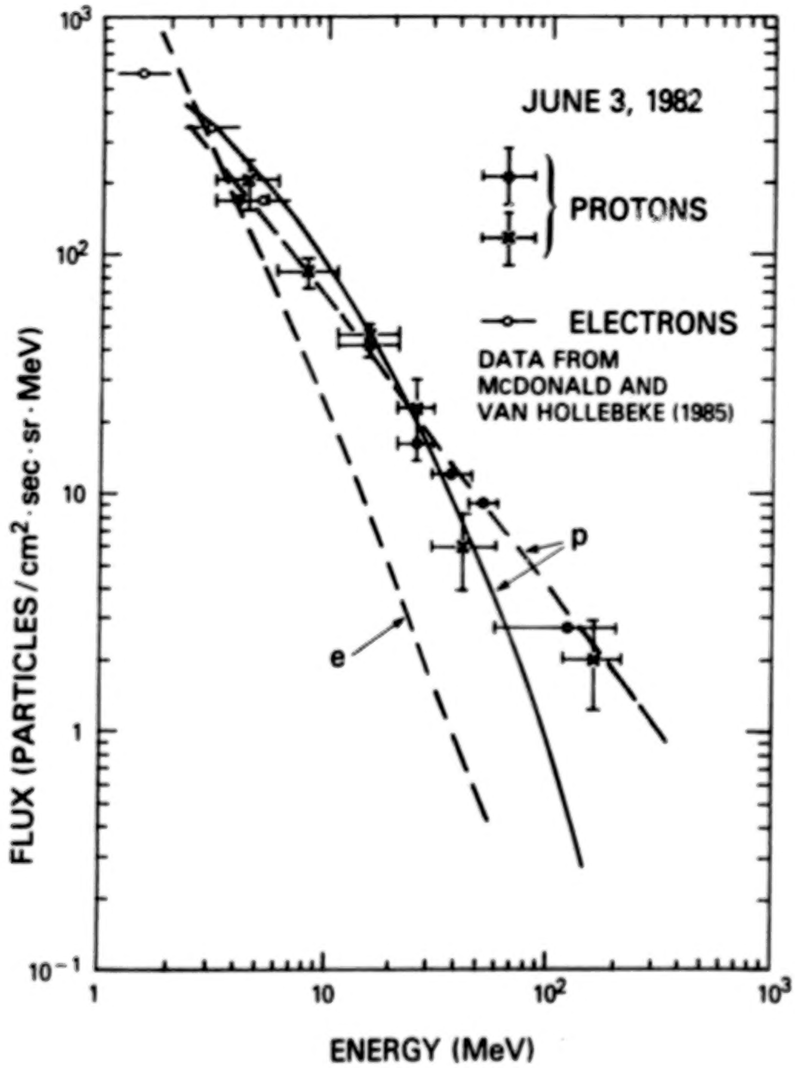


Figure 2. Interplanetary proton and electron observations [McDonald and Van Hollebeke, 1985] of the June 3, 1982 flare. The solid curve corresponds to a stochastic-acceleration spectrum with $\alpha T = 0.04$, and the dashed curves correspond to shock-acceleration spectra with $s = 2.4$ and E_{0p} and E_{0e} tending to infinity.

MeV lines, neutrons and interplanetary charged particles) can be explained by a model which incorporates two distinct particle populations with different acceleration histories and different but time independent energy spectra. They suggest that the softer proton spectrum is accelerated early in the flare by the first of two acceleration phases, while later in the flare second phase acceleration produces a much harder spectrum. We shall discuss these multiple acceleration phases in the next subsection, but before that we will comment on the maximum energy of particles accelerated in flares.

As just mentioned, the interplanetary proton spectrum of the June 3, 1982 flare was exceptionally hard, but these observations determined the proton spectrum only up to ~ 200 MeV. The ground-based neutron observations from this flare [Debrunner et al., 1983], which are very sensitive to protons of GeV energies, set an upper limit, $E_{0p} < 3$ GeV [Murphy, Dermer, and Ramaty, 1987]. The highest energy particles from flares have been seen with ground-based detectors. These show that the interplanetary proton spectrum from solar flares occasionally extends to energies > 10 GeV [Debrunner et al., 1984].

B. Trapping and Escape of Particles and Multiple Acceleration Phases

With a few exceptions (the flares of August 4, 1972 and June 3, 1982), the number of protons deduced from gamma ray and neutron observations exceeds by at least an order of magnitude the number obtained from interplanetary observations of the same flares. This implies that the gamma rays and neutrons are produced by particles accelerated in and confined to closed magnetic configurations, probably magnetic loops. On the other hand, there are many flares which produce large fluxes of interplanetary particles without producing detectable gamma rays or neutrons. The flare of December 9, 1981 is an example [Cliver et al., 1983]. The number of escaping particles from this flare exceeds the upper limit set [Vlahos et al., 1987] by gamma ray observation by a factor of 5. These particles are most likely accelerated at coronal sites with ready access to interplanetary space.

These results are consistent with the two phase acceleration model suggested originally by Wild, Smerd, and Weiss [1963]. Recently Cane, McGuire, and von Rosenvinge [1986] have related the two acceleration phases to two classes

of solar flare particle events which can be distinguished by a variety of signatures including the duration of their associated soft X-ray emission [Pallavicini, Serio, and Vaiana, 1977]. The first class, characterized by short duration impulsive events, is probably due to impulsive energy release in compact source regions relatively low in the corona. The second class, characterized by longer duration soft X-ray emission, corresponds to energy release in extended regions high in the corona. It is reasonable to associate the bulk of the gamma ray flares (including the June 7, 1980 and June 21, 1980 flares for which only a small fraction of the protons escaped from the Sun) with the first class or the first phase, and the interplanetary proton events which do not show detectable gamma rays (e.g., the December 9, 1981 flare) with the second class or phase. A similar classification has also been proposed recently by Bai [1986].

Of all flares, only the June 3, 1982 flare produced observable gamma ray emission which can be associated with second phase acceleration. As in other gamma ray flares, this flare also produced nuclear line emission and neutron capture radiation [Prince et al., 1983; Chupp et al., 1987]. These emissions are typical of first phase acceleration [Murphy, Dermer, and Ramaty, 1987]. But the June 3 flare also produced emission from the decay of pions [Forrest et al., 1987] and the time profile of the observed pion decay emission strongly suggests second phase acceleration [Ramaty, Murphy, and Dermer, 1987]. The number of escaping particles from this flare was smaller by about an order of magnitude than the number deduced from the gamma ray and neutron data in the first phase and exceeded by a factor of about 50 the number deduced from the pion decay gamma rays in the second phase. In addition, as discussed above, the spectrum of the escaping particles is much more consistent with the spectrum of the second phase particles than with that of the first phase ones. These results suggest that during its first phase the June 3 flare was similar to the other gamma ray flares in which the bulk of the interacting particles remained trapped and thermalized at the Sun. During the second phase of this flare, the bulk of the particles escaped. This phase of the June 3 flare resembles the flares which produce interplanetary particles without detectable gamma rays and neutrons. Second phase gamma rays were seen from the June 3 flare only because of its very hard proton spectrum which led to efficient pion production.

C. Energetic Particle Anisotropies in the Solar Atmosphere

The observations that continuum gamma ray emission is seen preferentially from flares close to the solar limb [Rieger et al., 1983; Vestrand et al., 1987] suggest that the angular distribution of the relativistic electrons in the region where the gamma rays are produced is anisotropic. It was suggested that this limb brightening could be due to an electron distribution that at injection was isotropic in the downward hemisphere [towards the photosphere, Petrosian, 1985], and by electron distributions in the interaction region which were peaked at directions either parallel or perpendicular (downward) to the photosphere [Dermer and Ramaty, 1986].

The angular distribution of the protons in the gamma ray production region can be studied by comparing the 2.223 MeV line, which is produced by neutrons moving downward into the photosphere, with the neutron flux observed near Earth. A recent study [Hua and Lingenfelter, 1987b] indicates that both an isotropic (4π) proton distribution and a distribution peaking at directions parallel to the photosphere are consistent with the data, but a distribution which is isotropic in only the downward hemisphere is not. This result, coupled with those on relativistic electrons, suggests that the gamma rays are probably produced by a mirroring distribution peaking at directions parallel to the photosphere.

D. Solar Elemental Abundances

The observation of energetic particles escaping from the Sun and gamma ray lines from solar flares have provided two new techniques for determining elemental abundances in the solar atmosphere. The elemental composition of the energetic particle depends on the composition of the ambient medium from which these particles are accelerated, but the injection and acceleration processes are expected to modify the composition significantly. Gamma ray line intensities are directly proportional to the abundance of elements in the ambient solar atmosphere, but so far the gamma ray spectrum of only one flare was analyzed in sufficient detail to determine abundances.

Nuclei heavier than He in solar energetic particles were first detected by Fichtel and Guss [1961] and since then many measurements of such particles have been made (see references in FRZ). The observed solar energetic particle composition is highly variable. The most dramatic departure of a solar energetic particle abundance from its photospheric value is that of ^{3}He [Garrard, Stone, and Vogt, 1973]. Here very large enhancements are observed in the ^{3}He/^{4}He ratio above its likely photospheric value [see Kocharov and Kocharov, 1984 for review]. These enhancements are probably caused by resonant heating of ^{3}He in the ambient atmosphere which could lead to the preferential accelerations of ^{3}He [Fisk, 1978; Kocharov and Kocharov, 1978]. It has been shown recently [Reames, von Rosenvinge, and Lin, 1985] that the preferential acceleration of ^{3}He is most likely a first phase phenomenon.

Observations of the 2.223 MeV line from solar flares have been used to determine the photospheric abundance of ^{3}He. It was shown by Wang and Ramaty [1974] that the photospheric ^{3}He is an important neutron sink and therefore the observed 2.223 MeV line, which results from neutron capture on ^{1}H, can set limits on the photospheric ^{3}He abundance. Recently, Hua and Lingenfelter [1987c] used the detailed June 3, 1982 observations of the 2.223 MeV line [Prince et al., 1983] to determine the ^{3}He abundance. They obtained $^3\mathrm{He/H} = (2.3 \pm 1.2) \times 10^{-5}$.

It would appear that the large variations of the observed composition of the solar energetic solar particles would preclude the determination of abundances in the ambient medium. Nonetheless, it has been suggested [Meyer, 1985a] that for non ^{3}He rich flares (for which the abundance variations are relatively small) the acceleration and injection effects could be separated from effects related to the ambient medium composition. In particular, it has been shown [Breneman and Stone, 1985] that the ratio of the abundance of elements observed in individual flares to the mean abundance (determined by averaging abundances of several flares) is a monotonic function, $(Q/M)^x$, where Q and M are ionic charge and mass, and x varies from flare to flare. These mean abundances could be similar to those of the ambient medium.

In Table 1 we show abundances derived from various sources. The local galactic (LG) set [Meyer, 1985b] is thought to represent photospheric abundances,

TABLE 1

Elemental Abundances

Element	Local Galactic[1]	Corona[1]	SEP[2]	Gamma Rays[3]
C	1260.00 (1.26)	600.00 (3.00)	271.00+-26.00	288.00+-50.00
N	225.00 (1.41)	100.00 (1.70)	77.50+- 5.20	117.00+-91.00
O	2250.00 (1.25)	630.00 (1.60)	623.00+-35.00	422.00+-62.00
Ne	325.00 (1.50)	90.00 (1.60)	88.70+- 8.70	199.00+-27.00
Na	5.50 (1.18)	7.00 (1.70)	7.33+- 0.70	- - - -
Mg	105.00 (1.03)	95.00 (1.30)	120.60+- 6.20	68.00+-25.00
Al	8.40 (1.05)	7.00 (1.70)	8.74+- 0.42	- - - -
Si	100.00 (1.03)	100.00 (1.30)	100.00	100.00+-28.00
P	0.94 (1.24)	- - - -	0.46+- 0.07	- - - -
S	43.00 (1.36)	22.00 (1.70)	22.20+- 0.75	48.00+-83.00
Cl	0.47 (1.60)	- - - -	0.21+- 0.07	- - - -
Ar	10.70 (1.50)	5.40 (1.80)	2.07+- 0.33	- - - -
K	0.34 (1.14)	- - - -	0.33+- 0.15	- - - -
Ca	6.20 (1.14)	7.50 (1.60)	6.80+- 1.10	17.00+-15.00
Ti	0.27 (1.16)	- - - -	0.38+- 0.11	- - - -
Cr	1.29 (1.10)	- - - -	1.43+- 0.27	
Mn	0.77 (1.24)	- - - -	0.52+- 0.25	
Fe	88.00 (1.07)	100.00 (1.50)	95.90+-10.00	76.00+-18.00
Ni	4.80 (1.13)	5.50 (1.70)	3.38+- 0.50	- - - -
Zn	0.10 (1.22)	- - - -	0.11+-0.05	- - - -

1. Meyer (1985b)

2. Breneman and Stone (1985)

3. Murphy et al. (1985)

even though the Ne abundance has not yet been measured in the photosphere. The coronal (COR) set, based on spectroscopic observations of the corona, is also from Meyer [1985b]. The mean solar energetic particle abundances (SEP) are from Breneman and Stone [1985]. The gamma ray (GR) set is from Murphy et al. [1985].

To compare these abundances, we have renormalized each pair of sets using multiplicative factors determined by minimizing χ^2. The ratios of elemental abundances for each pair are shown in Table 2; together with the number of degrees of freedom, n; the value of the reduced χ^2, χ_n^2; and the corresponding probability. Here the elements are ordered by decreasing first ionization potential. As can be seen, it is very improbable that the LG and SEP sets are drawn from the same underlying population. In particular, the SEP to LG ratios for elements with high ionization potential are lower than the ratios for elements with low ionization potential. A similar suppression has been found in the coronal abundances relative to the local galactic abundances (see the COR/LG ratios), although here the difference is much less significant. As we have argued above, and as suggested by charge state measurements [Gloeckler, 1985], the solar energetic particles are probably accelerated in the corona. The SEP abundance set could therefore represent coronal abundances. The differences between coronal and photospheric abundances could be caused by charge dependent mass transport from the photosphere to the corona. Since the photosphere is collisionally ionized at a relatively low temperature, the transport could be less efficient for elements with high ionization potential, leading to suppressed abundances for such elements.

The gamma ray set probably represents chromospheric abundances in a flare loop [Murphy and Ramaty, 1984]. As for the SEP/LG ratios, the GR/LG ratios for O and C are lower than the ratios for Mg, Si, and Fe. The ionization potentials of C and O are higher than those of Mg, Si, and Fe. The suppression of the C and O abundances could also be caused by charge dependent transport, from the photosphere to the chromosphere in this case. However, if the Ne abundance in the photosphere (where it cannot be measured) is the same as in the local galactic set, then the difference between the chromospheric and photospheric abundances must be due to additional processes, because correlation with first ionization potential alone would

TABLE 2

Ratios of elemental abundances for the six combinations of the data sets of Table 1. Also shown are the degrees of freedom, n, the minimal reduced X^2 and the corresponding probability, P(X), for each set of ratios. The elements are ordered by decreasing first ionization potentials.

Element	SEP/LG	GR/LG	SEP/COR	GR/COR	GR/SEP	COR/LG
Ne	0.27+-0.14	0.80+-0.41	0.93+-0.57	1.84+-1.13	2.17+-0.36	0.32+-0.25
Ar	0.19+-0.10	- - - -	0.36+-0.30	- - - -	- - - -	0.58+-0.55
N	0.34+-0.14	0.68+-0.59	0.73+-0.52	0.97+-1.02	1.46+-1.14	0.51+-0.41
O	0.27+-0.07	0.24+-0.07	0.93+-0.56	0.56+-0.34	0.66+-0.10	0.32+-0.21
Cl	0.42+-0.29	- - - -	- - - -	- - - -	- - - -	- - - -
C	0.21+-0.06	0.30+-0.09	0.43+-0.85	0.40+-0.80	1.03+-0.20	0.55+-1.10
P	0.48+-0.13	- - - -	- - - -	- - - -	- - - -	- - - -
S	0.50+-0.18	1.45+-2.56	0.95+-0.67	1.81+-3.39	2.09+-3.62	0.59+-0.46
Zn	1.05+-0.57	- - -	- - -	- - -	- - - -	- - - -
Si	0.97+-0.03	1.30+-0.37	0.95+-0.28	0.83+-0.34	0.97+-0.27	1.15+-0.35
Fe	1.06+-0.13	1.12+-0.28	0.91+-0.46	0.63+-0.35	0.77+-0.20	1.30+-0.66
Mg	1.12+-0.07	0.84+-0.31	1.20+-0.37	0.60+-0.28	0.55+-0.20	1.04+-0.31
Ni	0.69+-0.13	- - - -	0.58+-0.42	- - - -	- - - - -	1.32+-0.94
Mn	0.66+-0.35	- - - -	- - - -	- - - -	- - - - -	- - - -
Ti	1.37+-0.45	- - - -	- - - -	- - - -	- - - -	- - - -
Cr	1.08+-0.23	- - - -	- - - -	- - - -	- - - -	- - - -
Ca	1.07+-0.23	3.56+-3.18	0.86+-0.53	1.80+-2.01	2.42+-2.17	1.39+-0.86
Al	1.01+-0.07	- - - -	1.18+-0.83	- - - -	- - - -	0.96+-0.67
Na	1.30+-0.26	- - - -	0.99+-0.70	- - - -	- - - -	1.48+-1.06
K	0.94+-0.45	- - - -	- - - -	- - - -	- - - -	- - - -
n	19	8	12	8	8	12
X^2 (n)	2.29	2.47	0.15	0.51	3.61	0.85
P	$1.1x10^{-3}$	$1.1x10^{-2}$	>0.99	0.85	$<10^{-3}$	0.6

predict a lower Ne abundance than the one obtained from the gamma ray studies.

Also shown in Table 2 are ratios of the gamma ray set to the SEP and coronal sets, and the ratios of the coronal set to the local galactic set. The difference between the GR and SEP sets is significant. In addition to the different Ne abundances, we also note that Mg in the GR set is significantly lower than in the SEP set (see also Table 1). There is no simple interpretation for this result. Clearly, the effects of injection and acceleration on the mean SEP abundances are not fully understood. Also, gamma ray data from more flares are needed to establish the constancy or variability of the abundances derived from gamma ray spectroscopy.

Elemental compositions have also been studied for ^{3}He rich flares. Mason et al. [1986] found that the heavy ion enrichments seen in these flares are uncorrelated with the ^{3}He enrichment. As pointed out above, the ^{3}He enrichment is most likely due to preferential heating and acceleration. Mason et al. [1986] suggest that the heavy ion enrichments seen in ^{3}He flares are caused by enrichments in the ambient coronal composition.

4. SUMMARY

That particle acceleration plays a dominant role in the physics of solar flares has been known for some time. In the present paper we have emphasized phenomena related to the acceleration of ions and relativistic electrons. We treated the presently best understood acceleration mechanisms, stochastic acceleration, and shock acceleration. We presented recent charged particle, gamma ray, and neutron observations, and we discussed their implications. These observations are highly complementary. The gamma rays and neutrons provide information on the particles which thermalize in the solar atmosphere, while the charged particles observations provide a direct sample of the escaping particles. The combination of information on the escaping and trapped particle populations strongly suggests the existence of multiple acceleration phases, which could be the manifestation of multiple acceleration sites, times, and perhaps even mechanisms.

Particles accelerated in solar flares also provide information on elemental abundances in the solar atmosphere. This information is obtained by measuring the composition of the escaping particles, by observing the gamma ray lines which are produced from ambient nuclei excited by the accelerated particles, and by studying the 2.223 MeV line which gives information on the ^3He in the photosphere. Abundance variations in the chromosphere and corona are suggested by these observations.

REFERENCES

Achterberg, A., 1981, *Astron. and Astrophys.*, **97**, 259.

Achterberg, A., and Norman, C. A., 1980, *Astron. and Astrophys.*, **89**, 353.

Armstrong, T. P., Pesses, M. E., and Decker, R. B., 1985, in *Collisionless Shocks in the Heliosphere: Reviews of Current Research*, ed. B. T. Tsurutani and R. G. Stone (Washington, DC: AGU), p. 271.

Axford, W. I., 1982, in *Plasma Astrophysics*, ed. T. D. Tuyenne and T. Levy, ESA Publication SP-151.

Axford, W. I., Leer, E., and Skadron, G., 1977, *15th Internat. Cosmic Ray Conference Papers (Plovdiv)*, **11**, 132.

Bai, T., 1986, *Astrophys. J.*, **308**, 912.

Bai, T., Hudson, H. S., Pelling, R. M., Lin, R. P., Schwartz, R. A., and von Rosenvinge, T. T., 1983, *Astrophys. J.*, **267**, 433.

Breneman, H. H., and Stone, E. C., 1985, *Astrophys. J.*, **299**, L57.

Cane, H. V., McGuire, R. E., and von Rosenvinge, T. T., 1986, *Astrophys. J.*, **301**, 488.

Chupp, E. L. et al., 1982, *Astrophys. J.*, **263**, L95.

Chupp, E. L. et al., 1987, *Astrophys. J.*, in press.

Cliver, E. V., Forrest, D. J., McGuire, R. E., and von Rosenvinge, T. T., 1983, *18th Internat. Cosmic Ray Conference Papers (Bangalore)*, **10**, 342.

Debrunner, H., Fluckiger, E., Chupp, E. L., Forrest, D. J., 1983, *18th Internat. Cosmic Ray Conference Papers (Bangalore)*, **4**, 75.

Debrunner, H., Fluckiger, E., Lockwood, J. A., and McGuire, R. E., 1984, *J. Geophys. Res.*, **89**, 769.

Decker, R. B., Pesses, M. E., and Armstrong, T. P., 1981, *17th Internat. Cosmic Ray Conference Papers (Paris)*, **3**, 406.

Dennis, B. R., 1985, *Solar Phys.*, **100**, 465.

Dermer, D. C., and Ramaty, R., 1986, *Astrophys. J.*, **301**, 962.

Droge, W., and Schlickeiser, R., 1986, *Astrophys. J.*, **305**, 909.

Drury, L. O. C., Axford, W. I., and Summers, D., 1982, *Mon. Not. Roy. Astron. Soc.*, **198**, 833.

Eichler, D., 1979, *Astrophys. J.*, **229**, 419.

Ellison, D. C., 1984, *J. Geophys. Res.*, **90**, 29.

Ellison, D. C., and Ramaty, R., 1985, *Astrophys. J.*, **298**, 400.

Fermi, E., 1949, *Phys. Rev.*, **75**, 1169.

Fichtel, C. E., and Guss, D. E., 1961, *Phys. Rev. Letters*, **6**, 495.

Fisk, L. A., 1978, *Astrophys. J.*, **224**, 1048.

Forman, M. A., 1981, *Advances in Space Res.*, **1**, 41.

Forman, M. A., Ramaty, R., and Zweibel, E. G., 1986, in *Physics of the Sun*, ed. P. A. Sturrock (Dordrecht: D. Reidel Publishing Co.), p. 249.

Forman, M. A., and Webb, G. M., 1985, in *Collisionless Shocks in the Heliosphere: A Tutorial Review*, ed. R.G. Stone and B.T. Tsurutani (Washington, DC: AGU), p. 91.

Forrest, D. J., Vestrand, W. T., Chupp, E. L., Rieger, E., Cooper, J., and Share, G. H., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **4**, 249.

Forrest, D. J., Vestrand, W. T., Chupp, E. L., Rieger, E., Cooper, J., and Share, G. H., 1987, *Advances in Space Res. (COSPAR)*, in press.

Garrard, T. L., Stone, E. C., and Vogt, R. E., 1973 in *High Energy Phenomena on the Sun*, ed. R. Ramaty and R. G. Stone, NASA SP 342, p. 341.

Gloeckler, G., 1985, *Advances Space Res. (COSPAR)*, **4**, 127.

Goldman, M. V., and Smith, D. F., 1986, in *Physics of the Sun*, ed. P. A. Sturrock (Dordrecht: D. Reidel Publishing Co.), p. 325.

Holman, G. D., Ionson, J. A., and Scott, J. S., 1979, *Astrophys. J.*, **228**, 576.

Hua, X-M., and Lingenfelter, R. E., 1987a, *Solar Physics*, in press.

Hua, X-M., and Lingenfelter, R. E., 1987b, *Astrophys. J.*, (submitted).

Hua, X-M., and Lingenfelter, R. E., 1987c, *Astrophys. J.*, (submitted).

Hudson, H. S., 1985, *Solar Physics*, **100**, 515.

Kocharov, G. E., and Kocharov, L. G., 1978, *10th Leningrad Symp. on Cosmic Physics*, (Leningrad: A. F. Yoffe Physico-Technical Inst.), p. 38.

Kocharov, L. G., and Kocharov, G. E., 1984, *Space Sci. Rev.*, **38**, 89.

Kulsrud, R. M., and Ferrari, A., 1971, *Astrophys. and Space Sci.*, **12**, 302.

Lee, M. A., 1982, *J. Geophys. Res.*, **87**, 5063.

Lee, M. A., and Ryan, J. M., 1986, *Astrophys. J.*, **303**, 829.

Lin, R. P., 1985, *Solar Phys.*, **100**, 537.

Mason, G. M., Reames, D. V., Kleckler, B., Hovestadt, D., and von Rosenvinge, T. T., 1986, *Astrophys. J.*, **303**, 849.

McDonald, F. B., and Van Hollebeke, M. A. I., 1985, *Astrophys. J.*, **290**, L67.

McGuire, R. E., and von Rosenvinge, T. T., 1984, *Advances in Space Res. (COSPAR)*, **4**, No. 2-3, 117.

McGuire, R. E., von Rosenvinge, T. T., and McDonald, F. B., 1979, *16th Internat. Cosmic Ray Conference Papers (Kyoto)*, **5**, 61.

McGuire, R. E., von Rosenvinge, T. T., and McDonald, F. B., 1981, *17th Internat. Cosmic Ray Conference Papers (Paris)*, **3**, 65.

Melrose, D. B., 1974, *Solar Physics*, **37**, 353.

Melrose, D. B., 1980, *Plasma Astrophysics* (New York: Gordon and Breach).

Meyer, J. P., 1985a, *Astrophys. J. (Supp.)*, **57**, 151.

Meyer, J. P., 1985b, *Astrophys. J. (Supp.)*, **57**, 173.

Murphy, R. J., Dermer, C. D., and Ramaty, R., 1987, *Astrophys. J. (Supp.)*, in press.

Murphy, R. J., and Ramaty, R., 1984, *Advances Space Res. (COSPAR)*, **4**, No. 7, 127.

Murphy, R. J., Ramaty, R., Forrest, D. J., and Kozlovsky, B., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **4**, 2.

Pallavicini, R., Serio, S., and Vaiana, G. S., 1977, *Astrophys. J.*, **216**, 108.

Petrosian, V., 1985, *Astrophys. J.*, **299**, 987.

Prince, T. A., Forrest, D. J., Chupp, E. L., Kanbach, G., and Share, G. H., 1983, *18th Internat. Cosmic Ray Conference Papers (Bangalore)*, **4**, 79.

Ramaty, R., 1979, in *Particle Acceleration in Astrophysics*, ed. J. Arons et al. (New York: American Institute of Physics), p. 135.

Ramaty, R., 1986, in *Physics of the Sun*, ed. P. A. Sturrock (Dordrecht: D. Reidel Publishing Co.), p. 291.

Ramaty, R., Murphy, R. J., and Dermer, C. D., 1987, *Astrophys. J.*, (submitted).

Reames, D. V., von Rosenvinge, T. T., and Lin, R. P., 1985, *Astrophys. J.*, **292**, 716.

Rieger, E., Reppin, C., Kanbach, G., Forrest, D. J., Chupp, E. L., and Share, G. H., 1983, *18th Internat. Cosmic Ray Conference Papers (Bangalore)*, **10**, 338.

Skilling, J. A., 1975, *Mon. Not. Roy. Astron. Soc.*, **172**, 557.

Sonnerup, B. V. O., 1973, in *High Energy Phenomena on the Sun*, ed. R. Ramaty and R. G. Stone, NASA SP 342, p. 357.

Svestka, Z., and Fritzova, L., 1974, *Solar Physics*, **36**, 417.

Syrovatskii, S. I., 1981, *Ann. Rev. Astron. Astrophys.*, **19**, 163.

Toptygin, I. N., 1980, *Space Sci. Rev.*, **26**, 157.

Tsurutani, B. T., and Rodriquez, P., 1981, *J. Geophys. Res.*, **86**, 4319.

Uchida, Y., 1974, *Solar Physics*, **39**, 431.

Van Hollebeke, M. A. I., MaSung, L. S., and McDonald, F. B., 1975, *Solar Physics*, **41**, 189.

Vestrand, W. T., Forrest, D. J., Chupp, E. L., Rieger, E., Share, G. H., 1987, *Astrophys. J.* (submitted).

Vlahos, L. et al., 1987, in *Energetic Phenomena on the Sun*, ed. M. R. Kundu and B. Woodgate, NASA CP-2439, in press.

Wang, H. T., and Ramaty, R., 1974, *Solar Physics*, **36**, 129.

Wentzel, D. G., 1965, *J. Geophys. Res.*, **70**, 2716.

Wild, J. P., Smerd, S. F., and Weiss, A. A., 1963, *Ann. Rev. Astron. Astrophys.*, **1**, 291.

N87
24251
UNCLAS

4

THE ENERGY SOURCE OF THE INTERPLANETARY MEDIUM AND THE HELIOSPHERE

Eugene N. Parker
Enrico Fermi Institute and
Department of Physics
University of Chicago
Chicago, Illinois 60637

ABSTRACT

The activity of the interplanetary medium arises from occasional transient outbursts of the active corona and, for the most part, from the interaction of fast and slow streams in the solar wind. The basic driver is the heat input to the corona, both transient and steady. The fast streams issue from coronal holes where the heat input may be Alfvén waves with root mean squared (rms) fluid velocities of nearly 10^2 km/sec or may be wholly or in part the waves refracted into the hole from neighboring active regions. If the latter, then the character of the wind from the coronal hole depends upon the proximity and vigor of active regions, with significant differences between the polar and low latitude solar wind. In any case, there is no observational support for any of these ideas, so that the primary cause of the wind from the Sun, as well as any other similar star, is not without mystery. It is to be hoped that ground-based observations together with the new input from the Solar Optical Telescope and the International Solar Polar Mission may in time succeed in clearing up some of the basic questions.

1. INTRODUCTION

The interplanetary medium, and the entire heliosphere, are a consequence of the continual expansion of the solar corona. The activity of the interplanetary medium arises from the occasional transient heating of the corona, and the mixture of fast and slow regions of wind from different parts of the corona. The basic energy source is the occasional flare and the coronal heating that maintains the temperature of the corona. It is the purpose of this presentation to review what we know and do not know of the heating.

The expansion of the corona follows from its high temperature, and it is generally accepted that the high temperature is caused by the dissipation of motions initiated in the convective zone. The difficulty is that the motions, and their ultimate dissipation in the corona, have proved elusive.

The corona is conveniently classified into three distinct regions, depending upon the intensity of the X-ray emission. There is the active X-ray corona ($N = 10^{10}$ atoms/cm^3, $T = 2.5 \times 10^6$ K, $B = 10^2$ gauss), requiring an energy input of about 1×10^7 ergs/cm^2 sec [Withbroe and Noyes, 1977]. The gas is confined in the closed bipolar magnetic field, so that most of the energy goes into radiation and thermal conduction downward into the transient region. There is the quiet corona, emitting a faint glow of X-rays ($N = 10^8$ atoms/cm^3, $T = 1.5 \times 10^6$ K, $B = 10$ gauss) maintained by an energy input of 3×10^5 ergs/cm^2 sec). Finally there is the tenuous coronal hole, conspicuous by the absence of X-ray emission ($N = 0.5 \times 10^8$ atoms/cm^3, $T = 1.5 \times 10^6$ K, $B = 10$ gauss) requiring an energy input of about 1×10^6 ergs/cm^2 sec [Zirker, 1977; Withbroe and Noyes, 1977; Leer, Holzer, and Fla, 1982; and Withbroe et al., 1985]. The magnetic field of the coronal holes opens outward into space, permitting free expansion of the coronal gas, so that most of the energy goes into production of the solar wind.

The fast streams in the solar wind come from the coronal holes [Hundhausen, 1972; Krieger, Timothy, and Roelof, 1973; Zirker, 1977; and Rottman, Orrall, and Klimchuk, 1982]. Evidently most of the wind, including the slow streams, is produced in and around coronal holes, with the quiet regions contributing to the slow wind [see discussion in Withbroe and Noyes, 1977]. We

would expect that the coronal gas on any open field lines contributes to the wind. The lowest energy state for the magnetic field is a closed configuration, so the field is open only where it is sufficiently weak to be pushed out by the pressure of the coronal gas [Parker, 1963]. The division between the active and the quiet corona may depend as much upon the closure of the magnetic field as upon the field strength.

A close examination of the mechanisms available for heating the corona suggests that there are qualitative as well as quantitative differences between active coronal regions and coronal holes. First of all, Rosner, Tucker, and Vaiana [1978], have emphasized that the observations of the active corona, and the theoretical models constructed from those observations, make it clear that (a) the heat input is distributed along the emitting X-ray loops and (b) there is a direct relation between heat input (ergs/cm^2 sec) and magnetic field. The relationship depends very little on the dimensions L of the region [Golub et al., 1980]. Thus, the X-ray bright points (L $\sim$ 10^4 km) have approximately the same surface brightness as the X-ray corona above a normal active region (L $\sim$ 2×10^5 km). The evidence is that the regions of re-entrant, i.e., closed, field, are heated largely through the dynamical nonequilibrium of the wrapping and interweaving of the lines of force, whereas the only known mechanism for heating the coronal regions with open fields is the dissipation of hydromagnetic waves. Thus it seems that it must be the generation and dissipation of magnetohydrodynamic waves that largely produce the solar wind and the heliosphere. The problem is to confirm this general concept with concrete facts.

Observations show wave motions, but with such small amplitudes that they represent no more than 10^5 ergs/cm^2 sec [Athay and White, 1978, 1979a,b; Bruner, 1978]. Hence, if there are enough waves to heat the corona, the scale of the waves must be so small ($<10^4$ km) that they are not resolved in the spectroscopic studies [Cheng, Doschek, and Feldman, 1979; Feldman, 1983; and Habbal, Leer, and Holzer, 1984; see also the results for sunspots, Beckers, 1976; Beckers and Schneeberger, 1977]. Unresolved waves are part of the "microturbulence", contributing to the line widths which place an upper limit of about 25 km/sec on the rms velocity $<v^2>^{1/2}$. Sound waves of this amplitude carry negligible energy because the speed of sound is only about 2×10^2 km/sec. If we assume, then, that the microturbulence is entirely the

result of unresolved Alfvén waves all propagating upward along the field, the energy flux is bounded by the upper limit $\rho \langle v^2 \rangle V_A$, where V_A is the Alfvén speed with a value of the order of 2×10^3 km/sec throughout the entire corona. This upper bound is 2×10^7 ergs/cm^2 sec in the active corona but only 10^5 ergs/cm^2 sec in the coronal hole as a consequence of the low density. It is immediately evident that the Alfvén wave amplitude must be much larger, of the order of 75 km/sec, if the coronal hole is to be heated by wave dissipation. We may speculate that the gas density in the coronal hole is so small that its contribution to the observed line widths is negligible when integrated along the line of sight, so that the necessary 75 km/sec rms velocities are undetected. Such velocities are relatively small, in the sense that the ratio of the velocity amplitude to the phase velocity, and the fractional variation $\Delta B/B$ of the magnetic field, are small, approximately 0.05. So perhaps the coronal hole is heated by the dissipation of Alfvén waves, or perhaps fast mode waves, with periods of, say, 100 sec and rms velocities of the order of 50-100 km/sec. But this is only a conjecture.

There are difficulties of another type in the active corona. The upper limit on the total wave flux of 2×10^7 ergs/cm^2 sec is sufficient to supply the active corona, but most of it must be dissipated in the first pass up around the bipolar field. The downward wave flux at the foot points of each re-entrant line of force must not be more than a third of the upward wave flux if the net upward energy flux is to be 1×10^7 ergs/cm^2 sec. What is more, the dissipation must be equally effective over scales ranging from 10^4 km to 2×10^5 km, i.e. Alfvén transit times of 5 to 100 sec. Most of the power in the observed small-scale fluctuations in the Sun lies at periods of 100 sec or more, with no theoretical reason to expect much power at shorter periods. It must be remembered, too, that whatever the distribution of wave power over period, it all contributes directly to the upper limit of 25 km/sec on the rms velocity.

How, then, are we to imagine that the dissipation of Alfvén waves supplies 1×10^7 ergs/cm^2 sec more or less equally to all scales? If, without any theoretical or observational basis, we postulate waves of sufficiently short wavelength as to heat the ephemeral active regions, dissipating over distances of 2×10^4 km, then these same waves dissipating over 2×10^4 km would heat

only the lower ends of the normal coronal loops with scales of 2×10^5 km. But Rosner, Tucker, and Vaiana [1978] showed that the heating is broadly distributed, if not entirely uniform, along the X-ray coronal loops. If, on the other hand, we imagine that the small, medium, and large active coronal regions are each heated by waves at different frequencies which just happen to provide the same heat input at all scales from small to large, we violate the observational upper limit of 25 km/sec on the rms gas velocity. For even one such frequency band carrying 1×10^7 ergs/cm^2 sec provides the upper limit of 25 km/sec on the rms gas velocity if it is not dissipated, leaving no room for another band or two, one of which is dissipated to supply the necessary 1×10^7 ergs/cm^2 sec. To see how this works note that a velocity of 25 km/sec is necessary to transport 1×10^7 ergs/cm^2 sec upward along the field. If the waves are not dissipated (and no one knows how to dissipate waves of such small amplitude $\Delta B/B \sim 0.05$ in so short a distance), then they propagate undiminished up around the re-entrant field and back down into the photosphere at the other end. Hence, both ends of the field have upward and downward propagating waves with an rms fluid velocity of 25 km/sec and zero net energy flux. The observational limit of 25 km/sec permits no other waves. So there is no room for undissipated waves. Somehow, then, we would require that waves of small amplitude dissipate more or less *uniformly* and *completely* along fields with lengths anywhere from 10^4 km to 2×10^5 km. To achieve this requires physical effects unknown to this author.

This leaves us with the alternative that the active corona, enclosed in the re-entrant fields of bipolar regions on the surface of the Sun, is heated principally by the current sheets produced by the shuffling and intermixing of the footpoints of the field [Parker, 1979, 1982, 1983a,b, 1984, 1985; Low, 1985]. The dissipation is then the intrinsic dynamical nonequilibrium and continual neutral point reconnection in the current sheets. The input of 1×10^7 ergs/cm^2 sec follows from shuffling of the footpoints at the not implausible speeds of 0.5 km/sec, wrapping the individual flux tubes about their neighbors with pitch angles of the order of 10-20° [Parker, 1983c].

Unfortunately, at the present time there is no observational information available on either the oscillations of the footpoints of the fields (producing Alfvén waves, etc.) or the wandering of the footpoints among the neighboring footpoints (producing the dissipative current sheets). The individual

magnetic fibrils are not resolved in ground-based observations, so that their individual motions are simply not known. The determination of the motions of the fibrils is just one of the many fundamental tasks that awaits the Solar Optical Telescope (SOT) in the next decade. Without the SOT the basic motions of the fibrils, and the strain rate within the field, cannot be determined, and the power source of the corona will remain a matter of "not implausible" assumptions, i.e., ignorance. It should be pointed out that the high-speed turbulence and intense jets observed in the corona by Brueckner and Bartoe [1983] and Withbroe, Habbal, and Ronan [1985] may be a direct manifestation of the dynamical nonequilibrium of the current sheets in the corona.

What, then, of the coronal hole—the source of the high-speed streams in the interplanetary medium? Whatever shuffling and intermixing of the footpoints we may imagine, the associated strains propagate outward into interplanetary space at the Alfvén speed, so that the lines of force do not accumulate any significant mutual wrapping and interweaving. There is no formation of current sheets and no significant dissipation. And as already noted, there is no indication of sufficient microturbulence in coronal holes to provide the necessary 10^6 ergs/cm^2 sec, in the form of outward propagating Alfvén waves. Again we will have to turn to the Solar Optical Telescope to provide complete quantitative information on the oscillatory motions of the magnetic fibrils at the photosphere from which we might estimate the amplitude of the waves in the corona. It will be recalled that wave amplitudes (microturbulence) of 50-100 km/sec are required.

While waiting for the SOT we may hope that some ingenious ground-based observation, or more modest space observation, will provide preliminary clues. In the simplest case, Alfvén waves propagating along a slowly varying magnetic field and slowly varying fluid density have a velocity amplitude that varies as $\rho^{-1/4}$. The density decrease from the photosphere, where the number density is of the order of 10^{17} atoms/cm^3, to the coronal hole, where the density is perhaps 10^8 cm^3, is a factor of 10^9. Thus, 0.5 km/sec in the photosphere produces 100 km/sec in the coronal hole. In actual fact the abrupt decline of the gas density, the associated rapid expansion of the individual fibrils to

fill all available space, and the dynamical spicule phenomenon together make any quantitative extrapolation from the photosphere a more complicated operation.

Hollweg, Jackson, and Galloway [1982] have treated the propagation of Alfvén waves in coronal holes. More recently Davila [1985] has explored how the waves might boost the wind along to produce the high speed streams [Holzer and Leer, 1980; Leer and Holzer, 1980]. He does not discuss the origin of the Alfvén waves.

The role of spicules in coronal heating is an intriguing question [cf. Athay and Holzer, 1982; Withbroe, 1983; and Sterling and Hollweg, 1984], although it would appear that their effects do not extend more than 5×10^4 km above the transition region.

Seeking alternatives to direct supply of Alfvén waves to the coronal hole, Fla et al. [1984] have noted the possibility that fast mode waves are refracted into the hole from a neighboring active coronal region [Habbal, Leer, and Holzer, 1979]. The authors have provided a quantitative exposition of the phenomenon. The magnitude of the effect needs to be established in some way from observation. An obvious question is whether the vigor of the high-speed streams in the solar wind reflect the proximity of active regions to the coronal holes throughout the 11-year cycle of activity. Generally speaking, the coronal holes at low latitudes are closer to active regions than the polar coronal holes, so that a direct comparison of low altitude fast streams with the polar wind [cf. Orrall, Rottman, and Klimchuk, 1983] should be instructive. The International Solar Polar Mission (ISPM) will provide fundamental information on this question. Indeed the entire picture of the connection of high-speed streams at low latitudes to the polar coronal holes will be examined by the ISPM. At present we know only that the stream activity at the solar equator correlates best with the magnetic fields and coronal structure at latitudes $\pm 30°$ from the work of Wilcox and others [Wilcox, 1968; Svalgaard, Wilcox, and Duvall, 1974; Svalgaard and Wilcox, 1975; Svalgaard, et al., 1975; and Hundhausen, 1977].

Whatever the source of the waves, it is difficult to imagine heating coronal holes by any means other than the dissipation of magnetohydrodynamic waves.

The dissipation poses no evident problem. The coronal holes are heated gently over long distances ($10^{11} - 10^{12}$cm) out into space. Over such extended scales the Alfvén propagation times are 5×10^2 to 10^4 sec, i.e., 5 to 100 times the 100 sec wave period. There is sufficient time for a variety of dissipation effects to develop, e.g. nonlinear steepening [Hollweg, Jackson, and Galloway, 1982] phase mixing, Landau damping of the fast and slow modes [Barnes, 1966, 1969, 1974, 1979; Barnes, Hartle, and Bredekamp, 1971; Hung and Barnes, 1973 a, b, c; and Habbal and Leer, 1982]. A plane transverse Alfvén wave with small fluid motion $\vec{v}$ ($|\vec{v}| << V_A$) does not damp significantly. But if the Alfvén speed V_A varies in the transverse direction of the fluid motion $\vec{v}$ ($\vec{v} \cdot \nabla V_A \neq 0$), the wave becomes oblique, with a longitudinal component, which is then subject to damping. Indeed, any Alfvén wave of limited transverse scale has a longitudinal component which is subject to Landau damping. So the primary question appears to be the existence of sufficiently strong hydromagnetic waves, and the sources of such waves. Thus it is of primary importance to confirm or deny the existence of waves of 50-100 km/sec amplitude, which is a difficult, if not impossible, task because the coronal holes are so tenuous. Studies of the active and quiet coronas around the periphery of the coronal hole may be informative, in view of the ideas put forth by Fla et al. [1984].

In conclusion it seems to be that the heating of the coronal hole is not without mystery. The energy supply responsible for both the fast and slow streams is simply not clear. This is a fundamental gap in our understanding of the origin and the activity of interplanetary space in particular and stellar physics in general.

This work was supported by the National Aeronautics and Space Administration under grant NGL-14-001-001.

REFERENCES

Athay, R. G., and Holzer, T. E., 1982, *Astrophys. J.*, **255**, 743.

Athay, G., and White, O. R., 1978, *Astrophys. J.*, **266**, 1135.

Athay, G., and White, O. R., 1979a, *Astrophys. J.*, **229**, 1147.

ORIGINAL PAGE IS
OF POOR QUALITY

Athay, G., and White, O. R., 1979b, *Astrophys. J. Supp.*, **39**, 333.

Barnes, A., 1966, *Phys. Fluids*, **9**, 1483.

Barnes, A., 1969, *Astrophys. J.*, **155**, 311.

Barnes, A., 1974, *Adv. Electronics Electron Phys.*, **35**, 1.

Barnes, A., 1979, in *Solar System Plasma Physics*, Vol. I, ed. E. N. Parker, C. F. Kennel, and L. J. Lanzarotti (New York: North Holland), pp. 249-319.

Barnes, A., Hartle, R. E., and Bredekamp, J. H., 1971, *Astrophys. J. Letters*, **166**, L53.

Beckers, J. M., 1976, *Astrophys. J.*, **203**, 739.

Beckers, J. M., and Schneeberger, T. J., 1977, *Astrophys. J.*, **215**, 356.

Brueckner, G. E., and Bartoe, J. D. F., 1983, *Astrophys. J.*, **272**, 329.

Bruner, E. C., 1978, *Astrophys. J.*, **226**, 1140.

Cheng, C. C., Doschek, G. A., and Feldman, U., 1979, *Astrophys. J.*, **227**, 1037.

Davila, J. M., 1985, *Astrophys. J.*, **291**, 328.

Feldman, U., 1983, *Astrophys. J.*, **275**, 367.

Fla, T., Habbal, S. R., Holzer, T. E., and Leer, E., 1984, *Astrophys. J.*, **280**, 382.

Golub, L., Maxson, C., Rosner, R., Serio, S., and Vaiana, G. S., 1980, *Astrophys. J.*, **238**, 343.

Habbal, S. R., and Leer, E., 1982, *Astrophys. J.*, **253**, 318.

Habbal, S. R., Leer, E., and Holzer, T. E., 1979, *Solar Phys.*, **64**, 287.

Hollweg, J. V., Jackson, S., and Galloway, D., 1982, *Solar Phys.*, **75**, 35.

Holzer, T. E., and Leer, E., 1980, *J. Geophys. Res.*, **85**, 4665.

Hundhausen, A. J., 1972, *Coronal Expansion and the Solar Wind* (New York: Springer-Verlag).

Hundhausen, A. J., 1977, in *Coronal Holes and High Speed Wind Streams* (Boulder, CO: Colorado Associated University Press), pp. 292-319.

Hung, R. J., and Barnes, A., 1973a, *Astrophys. J.*, **180**, 253.

Hung, R. J., and Barnes, A., 1973b, *Astrophys. J.*, **180**, 271.

Hung, R. J., and Barnes, A., 1973c, *Astrophys. J.*, **181**, 183.

Krieger, A. S., Timothy, A. F., and Roelof, E. C., 1973, *Solar Phys.*, **29**, 505.

Leer, E., and Holzer, T. E., 1980, *J. Geophys. Res.*, **85**, 4681.

Leer, E., Holzer, T. E., and Fla, T., 1982, *Space Sci. Rev.*, **33**, 161.

Low, B. C., 1985, *Solar Phys.*, **100**, 309.

Orrall, F. Q., Rottman, G. J., and Klimchuk, J. A., 1983, *Astrophys. J. Letters*, **266**, L65.

Parker, E. N., 1963, *Interplanetary Dynamical Processes* (New York: John Wiley and Sons).

Parker, E. N., 1979, *Cosmical Magnetic Fields* (Oxford: Clarendon Press).

Parker, E. N., 1982, *Geophys. Astrophys. Fluid Dyn.*, **22**, 195.

Parker, E. N., 1983a, *Geophys. Astrophys. Fluid Dyn.*, **23**, 85.

Parker, E. N., 1983b, *Geophys. Astrophys. Fluid Dyn.*, **24**, 79.

Parker, E. N., 1983c, *Astrophys. J.*, **264**, 642.

Parker, E. N., 1984, in *Proceedings of III Trieste Workshop on Relations Between Chromospheric-Coronal Heating and Mass Loss in Stars, Sacramento Peak Observatory, 18-25 August*, ed. R. Stalio and J. B. Zirker, pp. 301-17.

Parker, E. N., 1985, *Geophys. Astrophys. Fluid Dyn.*, in press.

Rosner, R., Tucker, W. H., and Vaiana, G. S., 1978, *Astrophys. J.*, **220**, 643.

Rottman, G. J., Orrall, F. Q., and Klimchuk, J. A., 1982, *Astrophys. J.*, **260**, 326.

Sterling, A. C., and Hollweg, J. V., 1984, *Astrophys. J.*, **285**, 843.

Svalgaard, L., and Wilcox, J. M., 1975, *Solar Phys.*, **41**, 461.

Svalgaard, L., Wilcox, J. M., and Duvall, T. L., 1974, *Solar Phys.*, **37**, 157.

Svalgaard, L., Wilcox, J. M., Scherer, P. H., and Howard, R., 1975, *Solar Phys.*, **45**, 83.

Wilcox, J. M., 1968, *Space Sci. Rev.*, **8**, 258.

Withbroe, G. L., 1983, *Astrophys. J.*, **267**, 825.

Withbroe, G. L., Habbal, S. R., and Ronan, R., 1985, *Solar Phys.*, in press.

Withbroe, G. L., Kohl, J. L., Weiser, H., and Munro, R. H., 1985, *Astrophys. J.*, **297**, 324.

Withbroe, G. L., and Noyes, R. W., 1977, *Annual Rev. Astron. Astrophys.*, **15**, 363.

Zirker, J. B., 1977, in *Coronal Holes and High Speed Wind Stream*, ed. J. B. Zirker (Boulder, CO: Colorado Associated University Press).

N87
24252
UNCLAS

5

PARTICLE PROPAGATION CHANNELS IN THE SOLAR WIND

K. A. Anderson* and W. M. Dougherty
Space Sciences Laboratory,
University of California
Berkeley, California 94720

ABSTRACT

The intensities of low energy solar-interplanetary electrons and ions at 1 AU occasionally change in a "square wave" manner. The changes may be increases or decreases and they typically have durations of from one hour to a few hours. In some cases these channels are bounded by discontinuities in the interplanetary field and the plasma properties differ from the surrounding solar wind. In one case solar flare particles were confined to a channel of width 3×10^6 km at Earth. At the Sun this dimension extrapolates to about 12 000 km, a size comparable to small flares.

1. INTRODUCTION

Bartley et al. in 1966 reported that highly anisotropic fluxes of 1 to 13 MeV protons from solar flares occasionally changed their flow direction suddenly and by as much as several tens of degrees. McCracken and Ness [1966] showed that the change in flow direction was due to changes in the direction of the interplanetary magnetic field (IMF). The time for passage of these bundles of field lines over the spacecraft implied their widths were about 3×10^6 km at 1 AU. Domingo, Page, and Wenzel [1976] described sudden decreases in

*Also Physics Department

C-2

solar particle intensity lasting about one hour and emphasized that such an effect meant that little transport of particles across interplanetary field lines was occurring.

In 1968, Jokipii and Parker used Leighton's hypothesis of random walk of magnetic field lines associated with granules and supergranules [1964] to develop a picture of an interplanetary medium composed of a tangle of field lines frozen into the solar wind, but whose feet were carried about by the random motions at the solar surface. Jokipii and Parker noted that using a correlation length of 15 000 km—about the radius of a supergranule—the magnetic structure would be 3×10^6 km in size at 1 AU. This is close to the size of the filaments as determined by Bartley et al. and McCracken and Ness. These workers did not find changes in the solar particle intensity, anisotropy ratio or energy spectrum as the spacecraft entered the "filament". More recently the IMF has come to be regarded as containing many discontinuities, rather than being made up of many filaments [Burlaga, 1969].

In this paper we discuss changes in the intensity of low energy solar-interplanetary electron and ion intensities observed at distance 1 AU from the Sun on the International Sun-Earth Explorer ISEE-3 spacecraft. These changes are characterized by particle increases or decreases, often having a square wave appearance, and lasting from one to a few hours.

Our measurements extend the previous ones by showing that low energy electrons and ions are "channeled" and that the intensity in the narrow channels may be higher than in the surrounding IMF. The particle intensity changes are usually well-defined, occurring over distances $\lesssim$10,000 km. Most importantly, we show that the net particle flow may be quite different in these narrow channels compared to the flow just outside.

Electrons in the energy range 2 to 10 keV were measured by a swept electric field analyzer and ions of energy in the range 45 keV to a few MeV were measured by a pair of solid state detector telescopes, one of which was covered by a thin foil in order to achieve separation of electrons and ions.

Although the electrons we measured are of quite low energy (2 to 40 keV), such particles have very high speeds (9% to 35% the speed of light), but small

gyroradii. In a 5 nT magnetic field at 60° pitch angle the gyroradii are in the range 15 to 70 km. The proton gyroradii are on the order of several thousand kilometers. The axes of the detectors lie in the ecliptic plane and are swept over it by rotation of the spacecraft. The view angles of the electron detector are 10° in azimuth and ±23° in elevation. Because of the restricted viewing in azimuth, variations in the azimuth of the interplanetary magnetic field can cause apparent changes in the count rates, especially when the particle pitch angle distributions are highly anisotropic. We therefore have eliminated from our analysis events contaminated by large and rapid changes in the elevation angle of the interplanetary magnetic field.

2. OBSERVATIONS

Figure 1 illustrates several features of low energy ion and electron propagation channels. During this 24-hour interval two increases in electron intensity can be seen as well as a less well defined decrease beginning near 2000 UT (Universal Time). The ion intensity decreases in coincidence with the two electron increases but are not quite so sharply defined (the ion behavior is not shown in Figure 1).

The clear correlation with solar wind and interplanetary magnetic field parameters are quite striking in this example. The magnetic field changes at the beginning and end of the two electron intensity increases have been identified as tangential discontinuities [Tsurutani, private communication, 1985]. The magnetohydrodynamic (MHD) theory of such discontinuities requires the pressure on the two sides to be equal. In response the solar wind plasma density increases during the times the IMF has decreased.

We have surveyed 15 months of ISEE-3 data during 1978 and 1979 and find about 80 intensity changes, either increases or decreases, resembling those shown in Figure 1. These features are most frequent and clearest in the 2 to 10 keV electron measurements, although ion intensity changes often accompany the electron intensity changes. Only a few of the examples are so clearly defined as the ones in Figure 1 and fewer still are accompanied by tangential discontinuities. Our basic assumption on the nature of these characteristic particle intensity changes is that the particles are confined to a bundle of

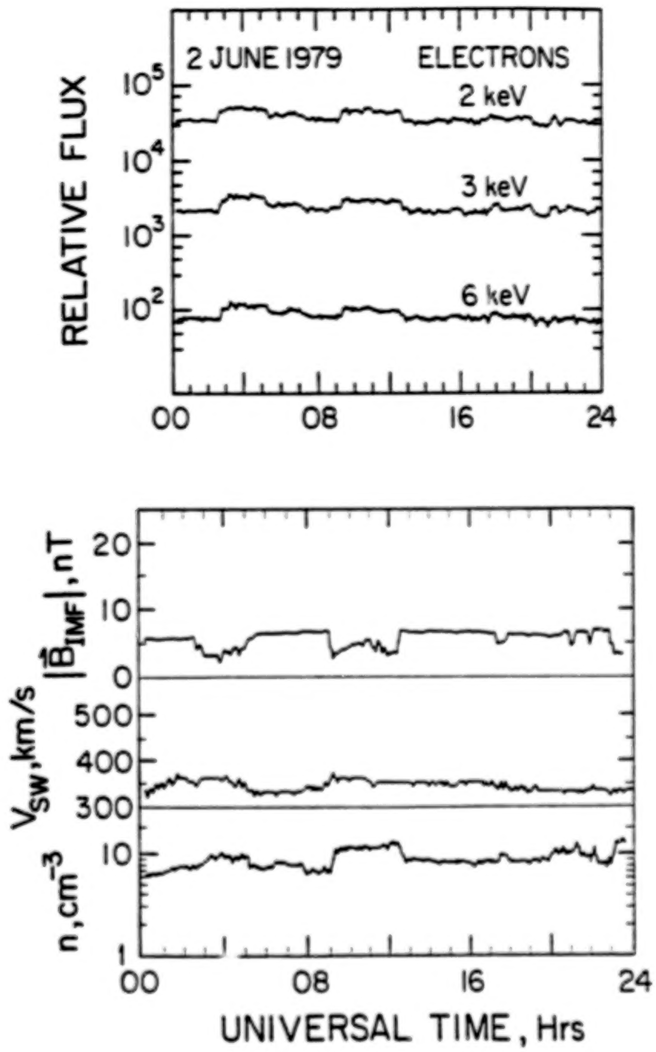


Figure 1. Changes in the intensity of low energy electrons and ions having a "square wave" appearance during long-lived streams of solar particles were a fairly common occurrence over 15 months of observations in 1978 and 1979. In the two-day period shown at least three such events occur. In some cases the intensity changes are associated with solar wind or interplanetary magnetic field changes. For some there is no clear signature.

magnetic field lines and that these field lines are swept past the spacecraft by the solar wind flow. We refer to these features in the IMF as particle propagation channels. Under this assumption we have converted the temporal duration of the propagation channels to a distance D given by

$$D = \sum_i V_{sw}^{(i)} \; |\sin \varnothing_i| \; \Delta t$$

where $V_{sw}^{(i)}$ is the solar wind speed averaged over time interval Δt, taken to be 10 seconds, and $\varnothing$ is the angle between the solar wind flow direction and the direction of the IMF during interval i.

We found the widths of the 80 propagation channels measured in this way to range from 1.4 to 5.0×10^6 km with an average value of 3.7×10^6 km.

McDonald and Burlaga [1985] have discussed another form of channeled solar particle propagation. They found regions of compressed interplanetary magnetic field evidently connecting back to the solar corona since solar particles were often found in them. Such channels are much larger in size than those we are discussing here. Their importance lies in the fact that adiabatic energy losses are reduced and thus the particles can reach interstellar space without large energy losses.

We next discuss an example which allows some conclusions about the origin and nature of one particular particle propagation channel. On May 20 and 21, 1979, several small solar flares occurred in McMath plage region 16014, located at about N16 W66. One of the flares in this region accelerated electrons from $\lesssim 2$ keV to above 40 keV and ions $\lesssim 40$ keV to above 270 keV energy (see Figure 2). The electron intensity versus time profile is characteristic of an impulsive injection of particles at the Sun into a medium whose mean free path for pitch angle scattering is on the order of 1 AU. The slow decay phase seen in Figure 2 is consistent with a brief injection phase followed by scattering in the interplanetary medium [Lin, 1974]. This view is supported by the measured pitch angle distributions. Initially, for some tens of minutes, the electrons are highly anisotropic, then become nearly isotropic.

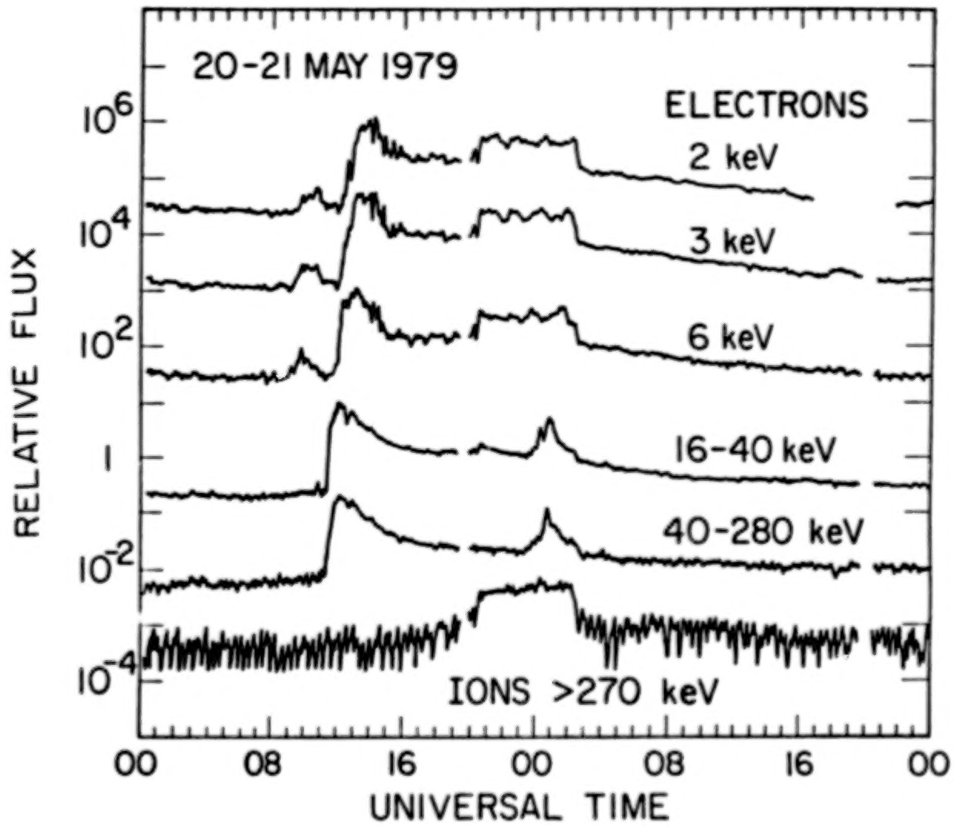


Figure 2. An importance 1N flare began at 1107 UT on May 20, 1979 and injected electrons from 2 to 100 keV and low energy ions into interplanetary space. The spacecraft entered a particle propagation channel at 2030 UT. This channel was 2.5 × 10^6 km in width at 1 AU.

The most remarkable feature during these two days is the sudden increase in the electron and ion intensities at about 2000 UT on May 20, followed about six hours later by a sudden return to the earlier slow decay. The increase is simultaneous at all particle energies. This is also the case for the decrease. During this six-hour period the pitch angle distributions became highly anisotropic showing strong flow of electrons away from the Sun (Figure 3).

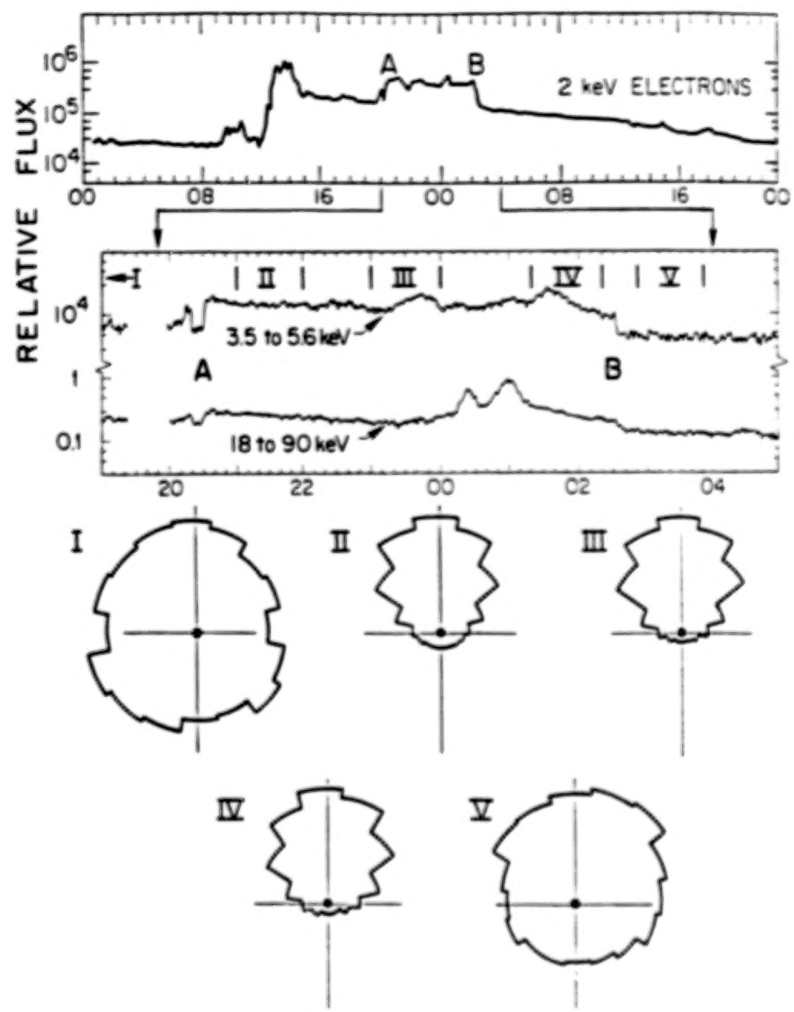


Figure 3. About nine hours after the solar flare that began at 1107 UT, highly anisotropic fluxes of electrons appeared for several hours (A to B). Before and after this time the electron fluxes were nearly isotropic. The vertical lines through the pitch angle distributions at the bottom of the figure lie along the interplanetary magnetic field. The direction of the Sun lies somewhere in the upper half of these diagrams. The pitch angle distributions are indexed to the middle panel by means of Roman numerals.

Evidently the interplanetary field lines have connected to a source region in the solar atmosphere able to continually inject electrons and ions into the interplanetary medium for at least six hours. We suppose that this injection also occurred before and after this six-hour interval. From the observed slow decay of particle intensity during this time, the source could have supplied particles for intervals as long as 20 hours or more.

From the solar wind speed and direction of the interplanetary magnetic field we find the width of the particle propagation channel to be 2.5×10^6 km. Linear extrapolation of this width to the solar surface gives a size of 12 000 km. The magnitude and topology of the magnetic field near the Sun no doubt will affect the actual size, but we appear to be dealing with a dimension which is larger than the size of a small flare but smaller than the plage region. Both at the Sun and in interplanetary space this structure is relatively small. One consequence of this is that if such structures are associated with most active regions they would be a common feature at 1 AU, but would not often be observed because of their small size.

While the spacecraft is located in the particle propagation channel three impulsive, flare-like injections of particles occur (Figure 3). They are identified by their impulsive appearance and their velocity dispersion, consistent with about a 1.3 AU travel distance. One of these electron events can be firmly identified with an optical flare in region 16014. No optical association could be found for the other two particle injections. This is not surprising since we have found from a study of many impulsive low energy electron events that those having very low intensity often have no reported Hα association [Potter, Lin, and Anderson, 1980]. However, it is unusual that three low energy solar electrons events occur in a six-hour interval. This leads us to believe that during the interval the spacecraft is situated on field lines connected to the site of flaring on the Sun. If this is the case the field lines guiding the electrons and ions away from the Sun from 2030 UT on May 20 to 0230 on May 21 have their origin in or near McMath plage region 16014, located at N16 W66 at this time. Since Earth is at a heliographic latitude of 2° S at this time the field lines would have been deflected 18° southward. This result is consistent with the suggestions of Schulz [1973] and by Svalgaard et al. [1975] and Svalgaard and Wilcox [1976], that the current sheet is related to the equator

of a solar magnetic dipole. In such a view field lines drawn out by the solar wind near the equator could trace back to latitudes well removed from the equator in the manner shown in Figure 7 of Smith, Tsurutani, and Rosenberg [1978].

The fact that energetic particles stream away from the Sun for periods of time on the order of days has been known for some time. [See, for example, Simnett, 1971 and Anderson, Lin, and Potter, 1982.] There are two general views on the mechanisms behind long-lived emission of solar particles. The first of these is storage of flare accelerated particles in the corona and their subsequent escape. The other hypothesis is continuous acceleration. Present observational evidence is not sufficient to resolve this issue, and there are conceptual difficulties with each hypothesis. Radio wave observations give the best evidence in favor of coronal storage, although the configurations of the magnetic fields have not been made clear in this way. The difficulty with coronal storage is the rapid rate of energy loss of the fast particles to background electrons in the coronal plasma. Only at very high coronal altitudes is the rate of energy loss low enough to permit storage over periods of days [Krimigis and Verzariu, 1971]. The problem with the continuous acceleration hypothesis is that no physical mechanism for it has been identified, and the best understood mechanisms involve shock waves and are therefore impulsive in character. Impulsive acceleration is noisy in the sense of copious X-ray and radio wave emission whereas these emissions are largely absent during periods of long-lived streaming. The present observations provide some additional information on the problem of long-lived emission of solar particles. In the first place only flares of small size are involved here whereas in the past the process has been generally associated with large flares. In the case at hand the largest flare was importance 1N and it accelerated only low intensities of ions in addition to the rather large fluxes of electrons. However, this flare occurred in an active region above which type III bursts frequently appeared, indicating that beams of fast electrons were present in the corona over a period of one or two days.

Secondly, the streaming of solar particles on May 20 to 21, 1979 is restricted to a spatial region whose dimension at the Sun is on the order of 10^4 km. The near perfect confinement of the streaming particles must be associated

with discontinuities or major changes in the topology of magnetic fields in the solar atmosphere. However, no magnetic feature could be identified with the particle channel on this occasion. In particular, nothing suggesting a neutral sheet appeared.

The third point we would make using data from the May 20 to 21, 1979 interval concerns the energy spectrum of the electrons. We noted above that the intensity of the 2 to 10 KeV electrons changed very little over a six-hour period. If the small change in intensity is interpreted as due to energy loss of these electrons to background electrons we can set a lower limit to the altitude at which the electrons must reside if indeed they are trapped in magnetic structures. A 10 to 20% energy loss in six hours at the lowest energies requires that the average electron density not exceed 10^5 cm^{-3}. This corresponds to a heliocentric distance of 4 solar radii for the quiet Sun and 19 solar radii if the coronal region is a streamer with enhanced density [Fainberg and Stone, 1971].

We have also compared the energy spectrum of the electrons in the impulsive injection from the 1N flare that began at 1107 UT on May 20 with the spectrum of electrons streaming in the propagation channel. The results are given in Table I. We have fitted the spectral data to a power law for 30 one-hour data samples and obtained the power law exponent for each sample. Ten samples are taken during times preceding entry into the particle channel, 10 samples in the channel, and 10 samples following exit from the channel. The average value for each set of ten samples and the standard deviations are given in Table I. We find that there is no difference in the low energy electron spectra between particles in the propagation channel and those on field lines outside the channel. There can be little doubt that the electrons outside the channel were accelerated by the 1N flare. We take the similarity of the energy spectra to be significant but not conclusive evidence for storage and subsequent escape over a period of many hours of electrons accelerated over a brief period of time by the importance 1N flare which began at 1107 UT on May 20, 1979. We conclude that some of the structure found in the interplanetary magnetic field is established by spatial features on the order of 10^4 km near the solar surface, and that these features persist for at least several days.

Table I.

Time 20-21 May 1979	Number of Intervals	Spectral Index, γ
14:00-19:00	10	3.77 ± 0.08
21:00-02:00	10	3.39 ± 0.29
03:00-08:00	20	3.49 ± 0.07

Comparison of the energy spectra of 2 to 10 keV electrons inside the particle propagation (2100 to 0200 UT) with spectra calculated before and after entry into the channel. The electrons outside the channel are certain to have been accelerated by an importance 1N flare, although at this time they are no longer streaming from the Sun. The close similarity of spectral indices suggests that the particles streaming away from the Sun in the propagation channel were accelerated by the same flare but then were trapped and released over a period of many hours.

ACKNOWLEDGMENTS

This work is the result of research supported by NASA grant NAG5-376.

REFERENCES

Anderson, K. A., Lin, R. P., and Potter, D. W., 1982, *Space Science Rev.*, **32**, 169.

Bartley, W. C., Bukata, R. P., McCracken, K. G., and Rao, U. R., 1966, *J. Geophys., Res.*, **71**, 3297.

Burlaga, L., 1969, *Solar Physics*, **7**, 54.

Domingo, V., Page, D. E., and Wenzel, K. P., 1976, *J. Geophys. Res.*, **81**, 43.

Fainberg, J., and Stone, R. G., 1971, *Solar Phys.*, **17**, 392.

Jokipii, J. R., and Parker, E., 1968, *Phys. Rev. Letters*, **21**, 44.

Krimigis, S. M., and Verzariu, P., 1971, *J. Geophys. Res.*, **76**, 792.

Leighton, R. B., 1964, *Astrophys. J.*, **140**, 1547.

Lin, R. P., 1974, *Space Sci. Rev.*, **16**, 189.

McCracken, K. G., and Ness, N. F., 1966, *J. Geophys. Res.*, **71**, 3315.

McDonald, F. B., and Burlaga, L. F., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **4**, 346.

Potter, D. W., Lin, R. P., and Anderson, K. A., 1980, *Astrophys. J.*, **236**, L97.

Schulz, M., 1973, *Astrophys. Space Sci.*, **24**, 371.

Simnett, G. M., 1971, *Solar Phys.*, **20**, 448.

Smith, E. J., Tsurutani, B. T., and Rosenberg, R. L., 1978, *J. Geophys. Res.*, **83**, 717.

Svalgaard, L., and Wilcox, J. M., 1976, *Nature*, **262**, 766.

Svalgaard, L., Wilcox, J. M., Scherrer, P. H., and Howard, R., 1975, *Solar Phys.*, **45**, 83.

Tsurutani, B. T., 1985, private communication.

N87
24253
UNCLAS

6

STUDIES OF THE INTERPLANETARY MAGNETIC FIELD: IMP'S TO VOYAGER

Norman F. Ness
Laboratory for Extraterrestrial Physics
Goddard Space Flight Center
Greenbelt, Maryland 20771

1. INTRODUCTION

It is my very great pleasure to participate in this special symposium honoring Dr. Frank McDonald and his significant contributions to space sciences. It was my good fortune in 1961 to become associated with Jim Heppner (Project Scientist and principal investigator) on the Explorer 10 project while at the NASA-Goddard Space Flight Center on an NAS-NRC postdoctoral Resident Research Associateship on a leave of absence from UCLA. At its termination in October 1961, I readily accepted an offer of permanent employment with NASA-GSFC, encouraged by both Frank McDonald and Les Meredith.

Following the 1961 successes of both the Explorer 10 and Frank's Explorer 12 spacecraft, Frank McDonald and colleagues proposed the first three spacecraft in a series to be called Interplanetary Monitoring Platforms or Probes. These were to be small, spin-stabilized, long-lived spacecraft placed in highly elliptical earth orbits to study the radiation environment in extraterrestrial space. It was proposed that I be the principal investigator for the Magnetic Field Studies on these IMP's and I was joined by Clell Scearce and Joseph Seek from the Explorer 10 team to round out the magnetic field instrument team.

The outstanding scientific successes of the 10 spacecraft which formed the IMP series is well known to almost all of you. It is not my intent to overview

all of those results here but only to highlight some of the main contributions related to the study of the Interplanetary Magnetic Field.

2. EARLY IMP PERIOD: 1963-1967

Earlier studies of the Interplanetary Magnetic Field by United States' spacecraft were of limited value due to spacecraft magnetic field contamination, incomplete vector measurements, limited data coverage, and limited spacecraft lifetime. With the full support of Frank McDonald, the technical staff at NASA-GSFC responded affirmatively to our request to double the length of the magnetometer booms on the IMP spacecraft from that originally planned. In addition, the other experiments and subsystem engineers cooperated fully in producing minimally magnetic electronic and detector modules so that the IMP spacecraft were the most magnetically clean spacecraft yet launched.

A brief summary of the IMP spacecraft program is presented in Figure 1 with launch date and Explorer designator indicated. The very significant increase in weight, for both spacecraft and experiments, for the last three in the series was due to the incorporation of solid strap-on rockets to create the thrust-augmented Delta (TAD) launch vehicle.

As the early IMP spacecraft were being constructed, tested, and integrated, Frank encouraged me to follow up on an idea of mine to place an IMP type spacecraft in close lunar orbit in order to study its magnetic field and radiation environment. Working with study manager Paul Marcotte and other staff at GSFC, we developed a proposal in late 1963 which subsequently led to the approval of the Anchored IMP series of two spacecraft for lunar studies. The chronology of that project is shown in Figure 2, illustrating the speed and vitality of the space program at that epoch: "The Swinging '60's".

One unique feature of the many spacecraft projects then being built at NASA-GSFC was the high esprit de corps which project staff and associated personnel maintained. This is illustrated in the cartoon shown in Figure 3, contrasting between the "cleanliness" of the Anchored IMP spacecraft procedures (due

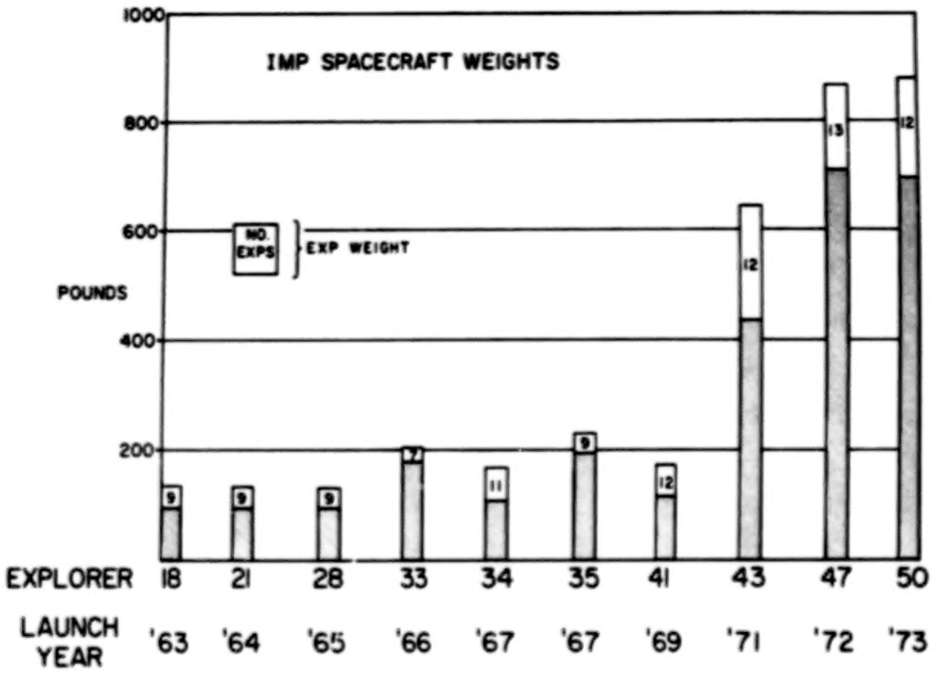


Figure 1. Summary of the Interplanetary Monitoring Platform spacecraft: launch date, Explorer numerical designation, spacecraft and experiment weight, and number of experiments.

to the Planetary Quarantine Requirements) and those of the regular IMP series. No one who participated in this period of space science shall ever forget the tremendous enthusiasm, dedication, and accomplishments which small, hard-working teams of scientists and engineers achieved in what appear now to be incredibly short periods of time.

The reputation and principal contributions of the Goddard Space Flight Center to the national space program have been in the area of customized spacecraft for scientific studies, and the IMP series under the tutelage of Frank McDonald best typifies that spirit.

EXPLORERS 33 & 35

AKA: LUNAR ANCHORED IMP/IMPs D & E

EVENT	DATE	APPROVAL
FINAL FEASIBILITY STUDY	NOVEMBER 27, 1963	H. GOETT
REVISED AO	DECEMBER 26, 1963	H. NEWELL
PROJECT APPROVED	JANUARY 20, 1964	H. NEWELL
PROPOSALS SUBMITTED	MARCH 1, 1964	---
EXPERIMENTS SELECTED	AUGUST , 1964	H. NEWELL
EXPLORER 33/IMP D LAUNCHED	JULY 1, 1966	TAD #39
EXPLORER 35/IMP E LAUNCHED	JULY 19, 1967	TAD #50

Figure 2. Chronology of the Anchored IMP mission, from proposal to launch. Approval by cognizant officials indicated. Dr. Harry Goett, then Director of the Goddard Space Flight Center and Dr. Homer Newell, Associate Administrator for Space Sciences of NASA.

3. SECTOR STRUCTURE OF INTERPLANETARY MAGNETIC FIELD

The data obtained by the IMP's 1, 2, 3, and 4 provided the first definitive measurements of the Interplanetary Magnetic Field structure and its variations. Figure 4 presents the distribution function of the magnitude of the Interplanetary Magnetic Field, as measured initially by the IMP-1 spacecraft

ORIGINAL PAGE IS
OF POOR QUALITY

Figure 3. Cartoon typifying the friendly but competitive spirit between the regular IMP and anchored IMP projects. Due to planetary quarantine restrictions, the AIMP spacecraft were constructed and integrated under substantially more restrictive conditions than the regular IMP series.

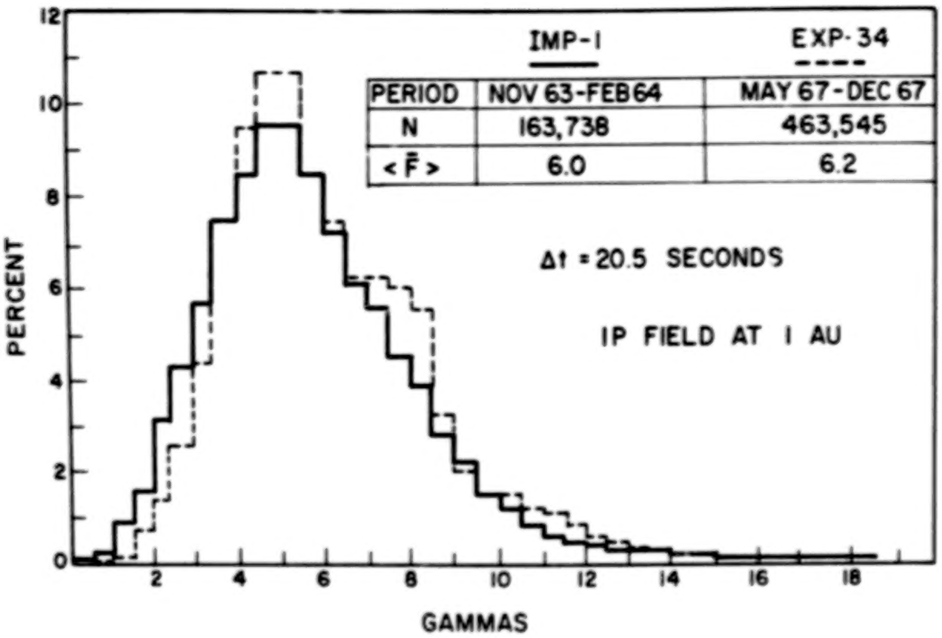


Figure 4. Histogram of the instantaneous magnitude of the interplanetary magnetic field as measured by the IMP-1 and IMP-4 spacecraft during the time those spacecraft were in the interplanetary medium outside the Earth's bow shock. Note that there is no substantial difference in the distribution, although the average field intensity is slightly larger for the later time interval.

in 1963-1964 and the IMP-4 spacecraft in 1967. In addition to confirming the average Archimedian spiral structure of the Interplanetary Magnetic Field, these early results showed an ordering of the polarity of the magnetic field which my colleague John Wilcox and I termed the interplanetary sector structure. (See Figure 5.) We were able to correlate these observations with the solar magnetic field and thereby establish conclusively its solar origin.

Nearly continuous observations of the Interplanetary Magnetic Field and sector structure was possible during the 1960's and 1970's as a result, primarily, of the data from the IMP series of spacecraft. More recent results in 1970 are shown in Figure 6, with an overlay on the planetary magnetic activity index

ORIGINAL PAGE IS OF POOR QUALITY

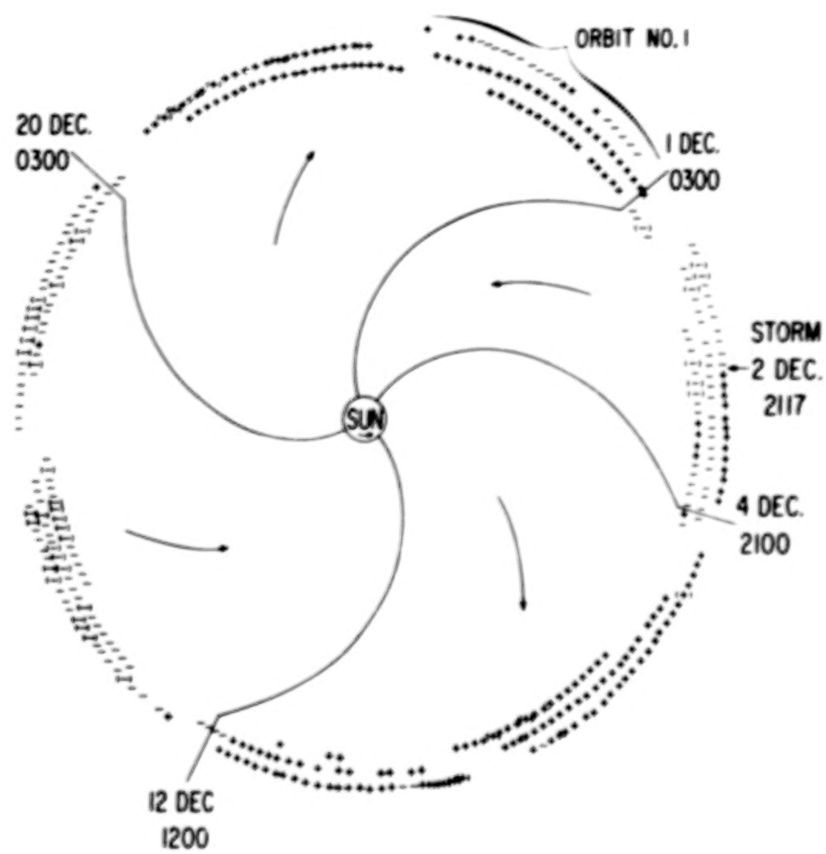


Figure 5. A well-known figure illustrating the sector structure of the interplanetary medium discovered by the IMP-1 spacecraft and measured over three solar rotations. [Ness and Wilcox, 1965]

(K_p) plots. Note here that only two sectors are shown in contrast to the earlier observations of four sectors. Also note that the location of the sector boundaries in heliographic longitude is time variable throughout this interval.

In addition to observations of the interplanetary sector structure at 1 AU, spacecraft enroute to encounters with the planets showed that the sector structure extended throughout interplanetary space. Observations by the Mariner

ORIGINAL PAGE IS
OF POOR QUALITY

102
Dec 70

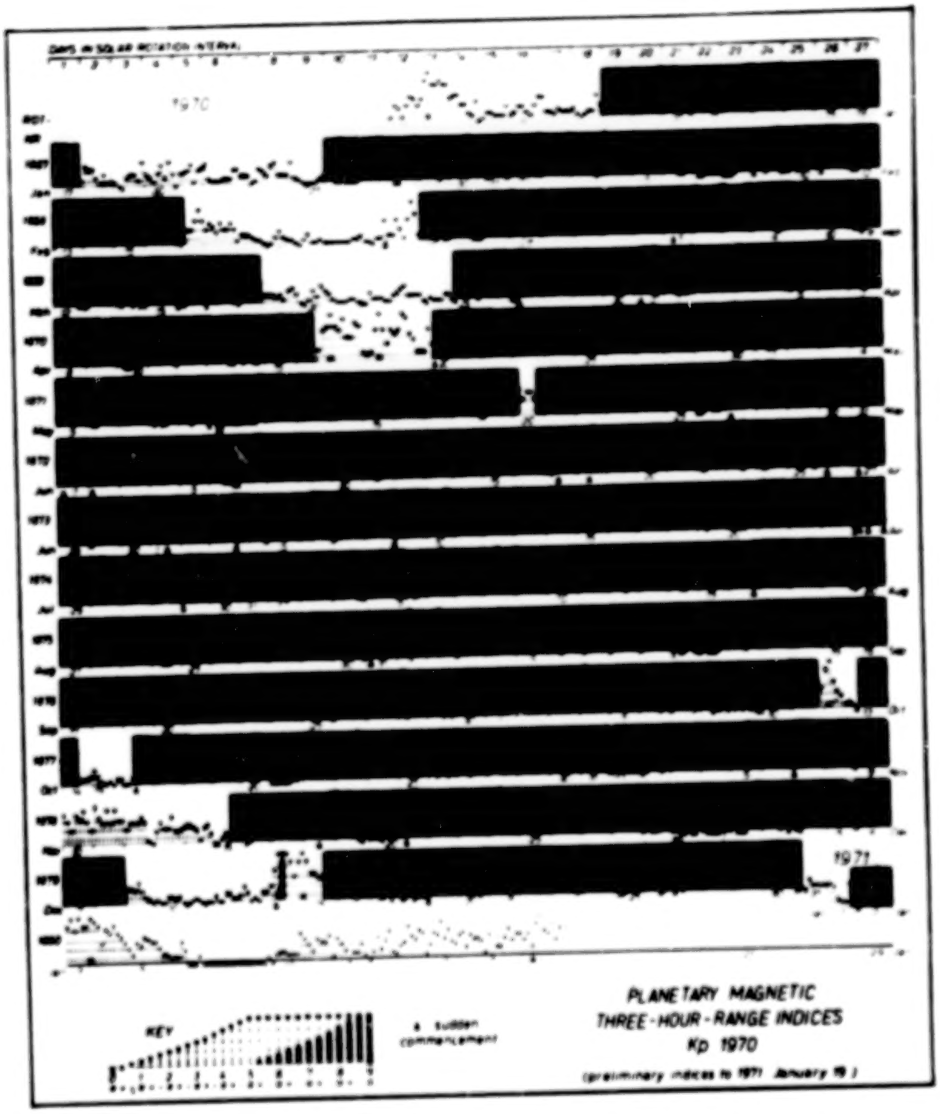


Figure 6. Overlay of the interplanetary sector structure during 1970 on the planetary magnetic activity index K_p. Note that only two sectors are shown and that the structure is not stationary during this time interval.

10 spacecraft (studied by Ken Behannon in his Ph.D. thesis) are shown in Figure 7. Again, a two-sector structure is observed throughout the several solar rotations covered by this data.

The understanding of the change from two to four sectors and the evolution of the sector structure boundaries with time initially eluded a clear explanation. That is until the concept of the heliospheric current sheet, analogous to that in Earth's magnetic tail separating regions of opposite magnetic polarity, was developed theoretically and empirically by a number of authors [Schulz, 1973; Rosenberg et al., 1973].

The inclination of the solar dipolar magnetic field axis to the rotation axis of the Sun will, together with a radial solar wind flow, transport the solar field into interplanetary space and lead to a two sector structure as illustrated in Figure 8. More complicated, i.e., a curved surface rather than a plane, configurations of the interplanetary neutral sheet arise since the solar magnetic field is not often well represented by a pure dipole. This is shown in Figure 9 where a change of sector structure from two to four sectors observed in the ecliptic at one AU occurs during three successive solar rotations.

The origin and location of the interplanetary neutral sheet on the Sun has been closely studied for sometime with observations obtained by the K coronameter brightness data. Figure 10 presents a plot of the contours of this parameter on the solar disk, with a superposition of the polarity of the magnetic field observed by the Voyager 1 and 2 spacecraft. It is seen that the correlation with the minimum brightness is excellent.Thus, the continuing observation of the interplanetary medium by the IMP and Voyager spacecraft has provided a substantial data set, not only for the study of the interplanetary medium, but for elaborate investigations of the response of the terrestrial magnetosphere to variations in the interplanetary medium structure.

4. COSMIC RAY MODULATION

One of the fundamental problems of cosmic ray studies has been to identify the mechanism responsible for both short- and long-term modulation of the

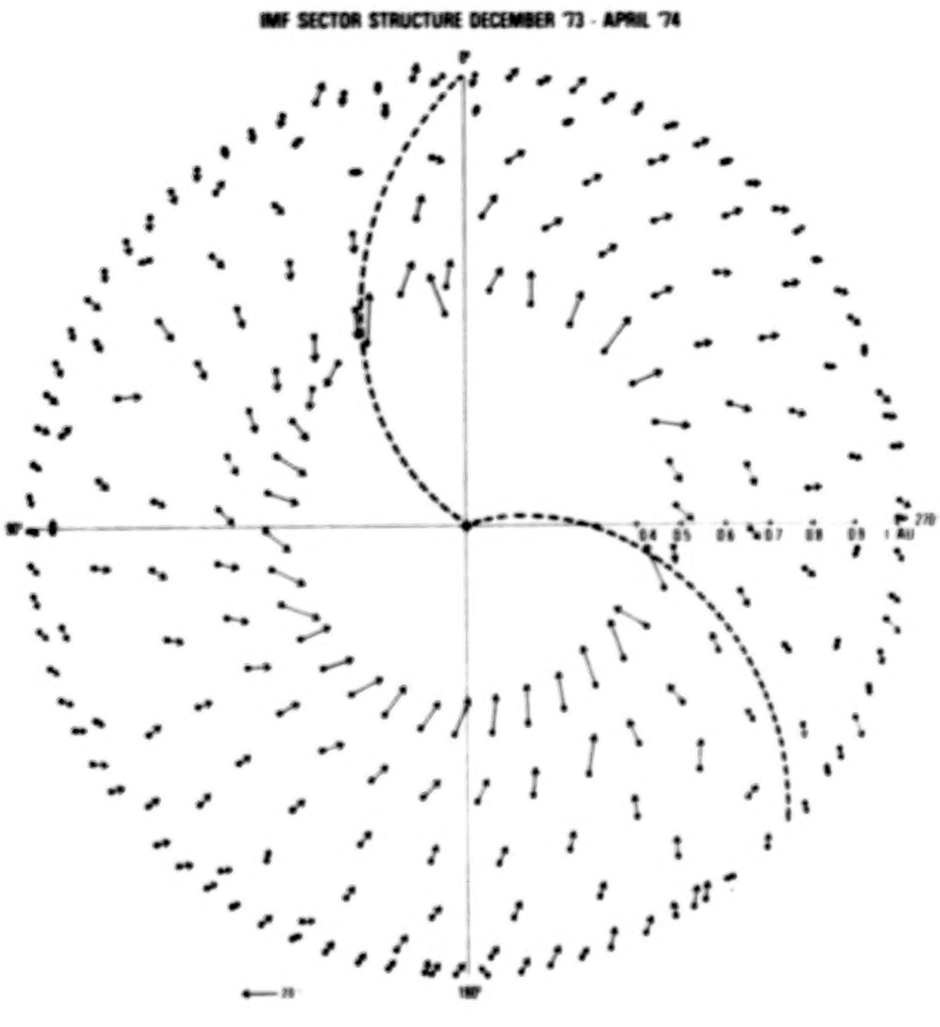


Figure 7. Interplanetary Magnetic Field sector observations obtained by the Mariner-10 spacecraft as it passed from 1 AU to encounter with Mercury in March 1974. The direction and intensity of the interplanetary field over one day intervals are shown by their vector projection. [Behannon, 1976]

ORIGINAL PAGE IS
OF POOR QUALITY

Figure 8. Sketch of the orientation of the heliospheric neutral sheet and the extended solar dipolar magnetic field, leading to a two-sector structure in the interplanetary medium. The polarity in the ecliptic depends upon whether the observations are made above or below the equatorial plane.

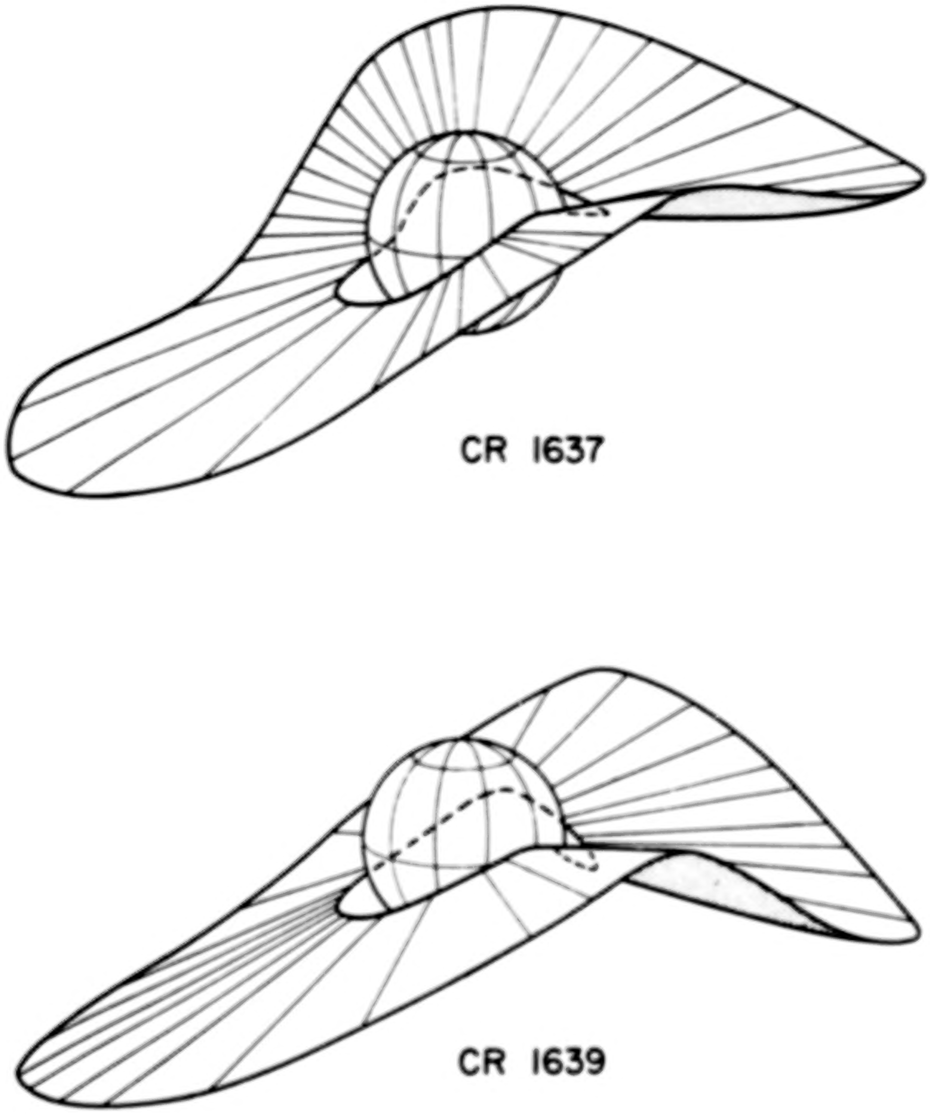


Figure 9. Sketch of the configuration of the heliospheric neutral sheet for Carrington solar rotations 1637 and 1639. During this time interval the sector structure changed from 2 to 4 sectors, as observed at 1 AU in the ecliptic. [Burlaga, Hundhausen, and Zhao, 1981]

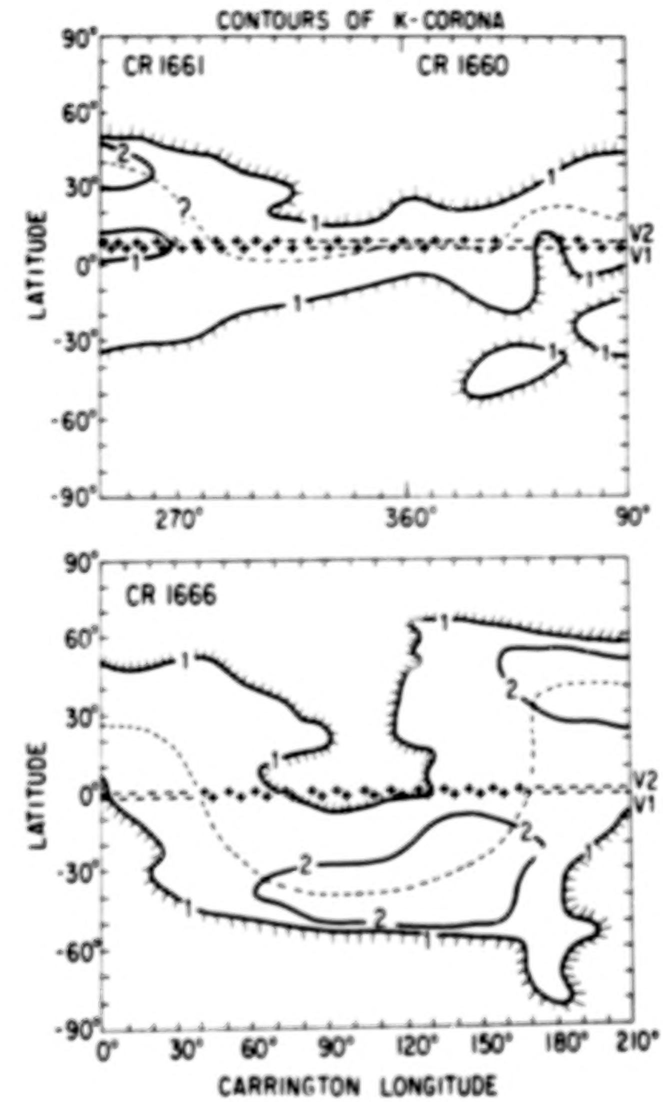


Figure 10. Plots of the K corona-meter brightness contours, used to investigate the configuration of the interplanetary neutral sheet at its source location on the Sun. Superimposed are the traces of the footprint of the polarities observed by the Voyager 1 and 2 spacecraft shortly after launch in 1977. [Behannon, Burlaga, and Hundhausen, 1983]

observed cosmic ray flux. The IMP series of spacecraft and other experiments of Frank McDonald on other spacecraft such as HELIOS, Pioneer and Voyager have contributed significantly to a resolution of some of the major issues related to this problem.

Different concepts of the magnetic structure of interplanetary space and the magnetic field configuration responsible for cosmic ray modulation are illustrated in Figure 11. Most of these were proposed prior to the recognition that interplanetary space is filled continually with the solar wind flux transporting solar magnetic fields into interplanetary space. With the knowledge of a continual solar wind flux, the question arose as to how variable-velocity solar plasma streams (or jets) would evolve in interplanetary space.

Figure 12 illustrates qualitatively the result of a high-speed plasma stream overtaking one of lower velocity. These "co-rotating" stream-stream interactions are a characteristic feature of the large-scale structure of the interplanetary medium, and a study of their radial evolution has been made possible with observations on solar and planetary probes.

Simultaneous observations of the interplanetary medium by the IMP-7 spacecraft at 1 AU and the Pioneer 10 spacecraft at 5 AU, just prior to encounter with Jupiter in 1973, demonstrate the changes which occur as a high-speed stream moves to 5 AU (see Figure 13). The concept of "filtering" of isolated streams and short wavelength speed fluctuations with streams is an important one. It has only recently been proposed as a result of data obtained and analyzed from the constellation of Pioneer, IMP, HELIOS, and Voyager spacecraft distributed throughout the heliosphere.

On a very large scale, Figure 14 shows how two co-rotating solar wind streams on opposite sides of the Sun with lifetimes of many solar rotations would modify the structure of the heliosphere. It is the compression region with enhanced magnetic fields which are found to be responsible for modulation of the cosmic ray flux observed both terrestrially and on deep space probes.

Correlation of the solar wind speed, Interplanetary Magnetic Field, and cosmic ray proton flux observed in the inner solar system is shown in Figure 15. Here

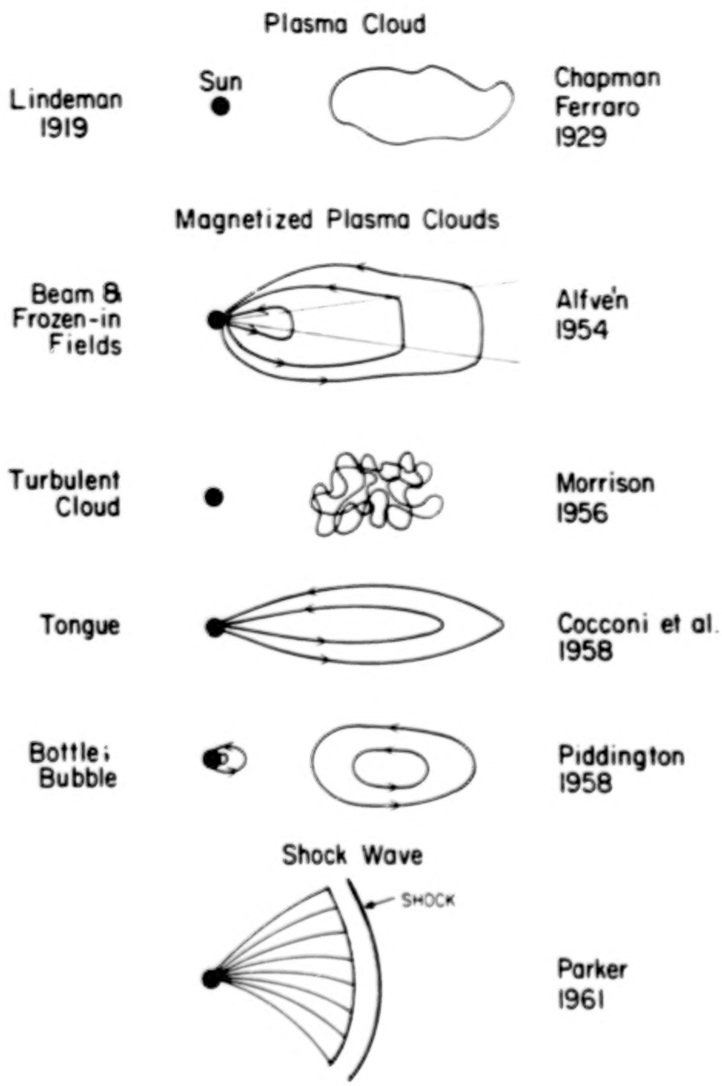


Figure 11. A summary of different concepts of the magnetic structure of interplanetary space considered responsible for cosmic ray modulation. [Burlaga, 1983]

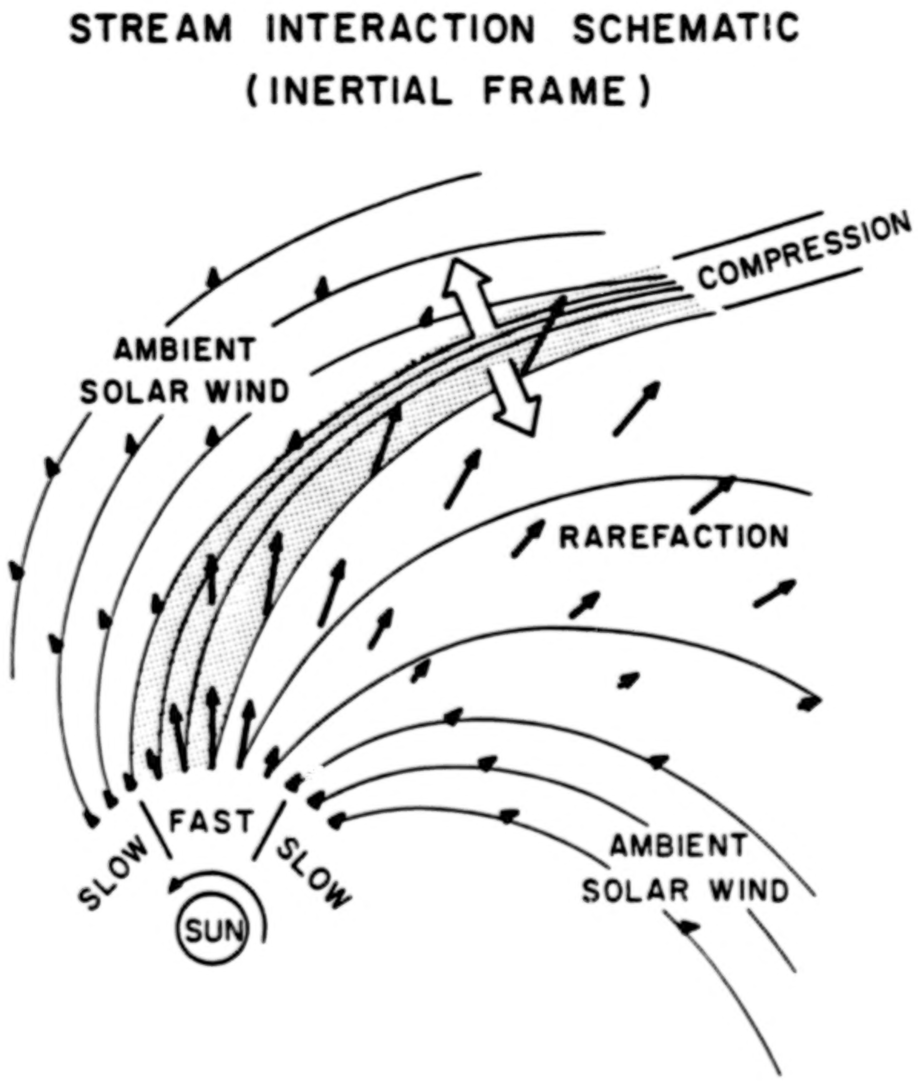


Figure 12. Diagram of the evolution, in interplanetary space, of the evolution of a co-rotating stream and interaction region. [Pizzo, 1978]

ORIGINAL PAGE IS
OF POOR QUALITY

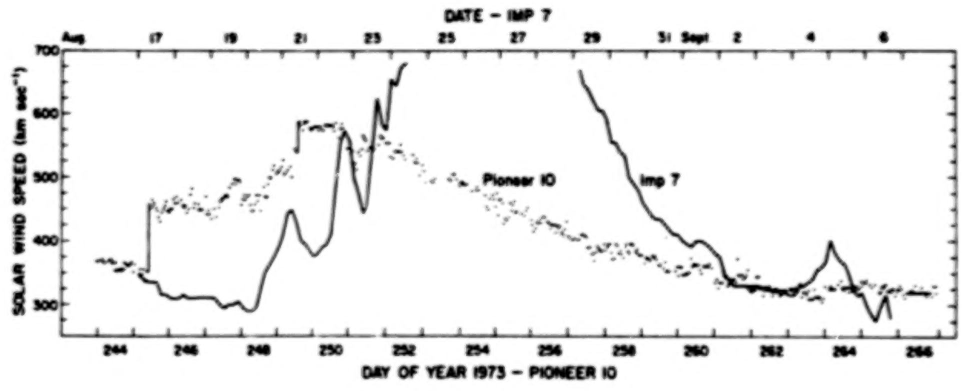


Figure 13. Comparison of solar wind speed profiles observed by IMP-7 at 1 AU and Pioneer-10 at 4.65 AU, illustrating the filtering or damping of large amplitude, short wavelength speed fluctuations at larger radial distances. [Gosling et al., 1976]

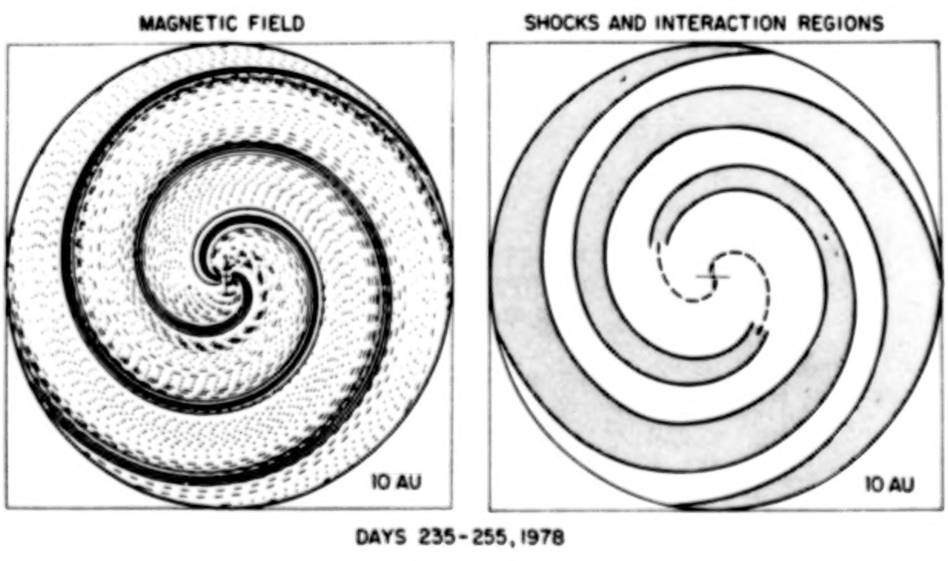


Figure 14. Diagram illustrating the evolution of the magnetic field geometry and compression regions as well as shocks and interaction regions associated with two co-rotating, high-speed solar wind streams on opposite sides of the Sun. [Burlaga, 1984]

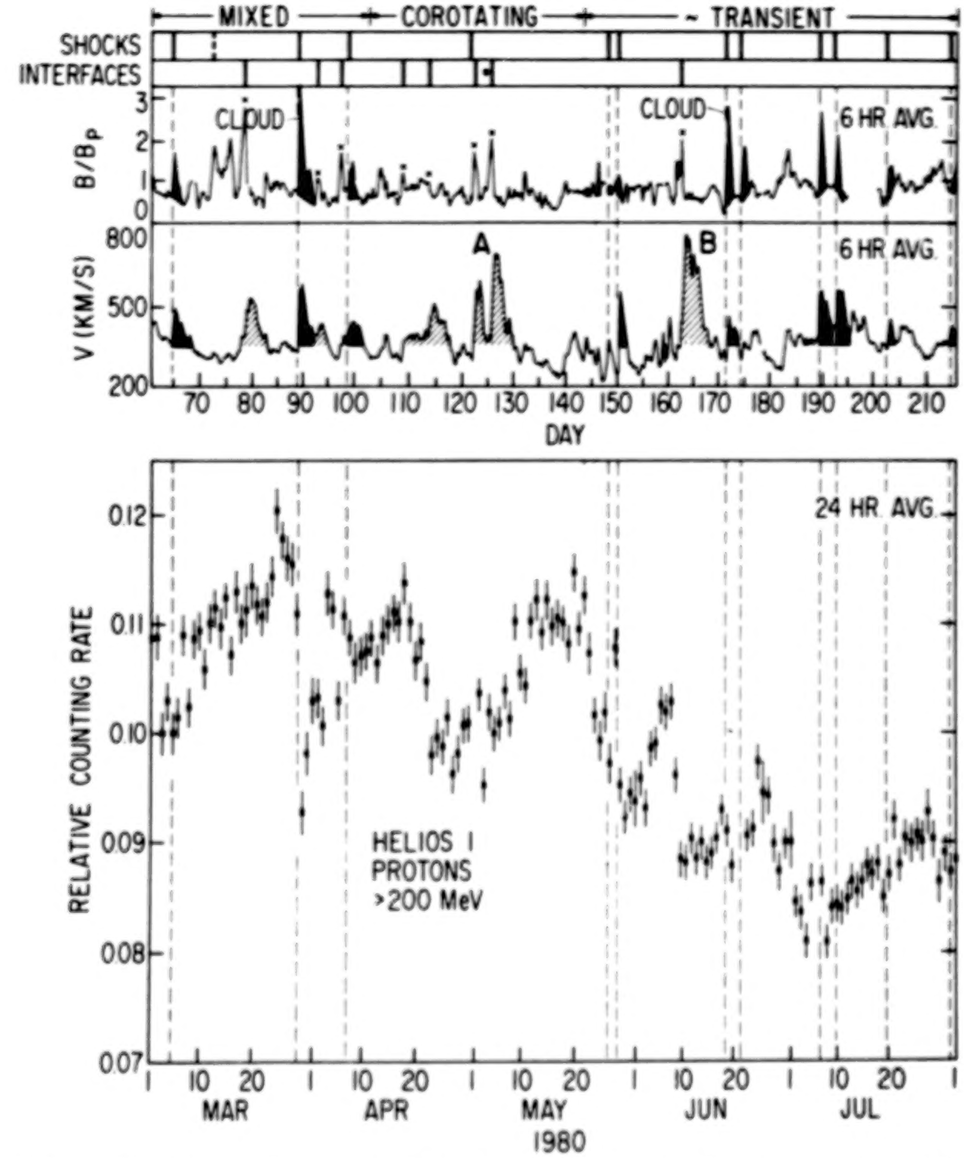


Figure 15. Observations of the flux of protons in interplanetary space as observed by HELIOS-1 compared with the solar wind and Interplanetary Magnetic Field parameters, emphasizing the modulation by co-rotating and transient features in the heliospheric structure. [Burlaga et al., 1984]

the important features are the modulation of the cosmic ray flux by the co-rotating, as well as transient, features in the interplanetary medium. Similar correlations have been observed at much greater heliocentric distances, and these are illustrated in Figure 16. Again, it is the compression region of significantly enhanced magnetic fields that are seen to be associated with the sudden decrease in the cosmic ray flux. It is known that the level of turbulence in the Interplanetary Magnetic Field is higher in the compression regions than in the rarefaction regions.

Frank McDonald, working in close collaboration with Len Burlaga, has conducted a systematic study of the time variations of cosmic ray flux, and solar wind structure, both in the inner and outer regions of the solar system. These

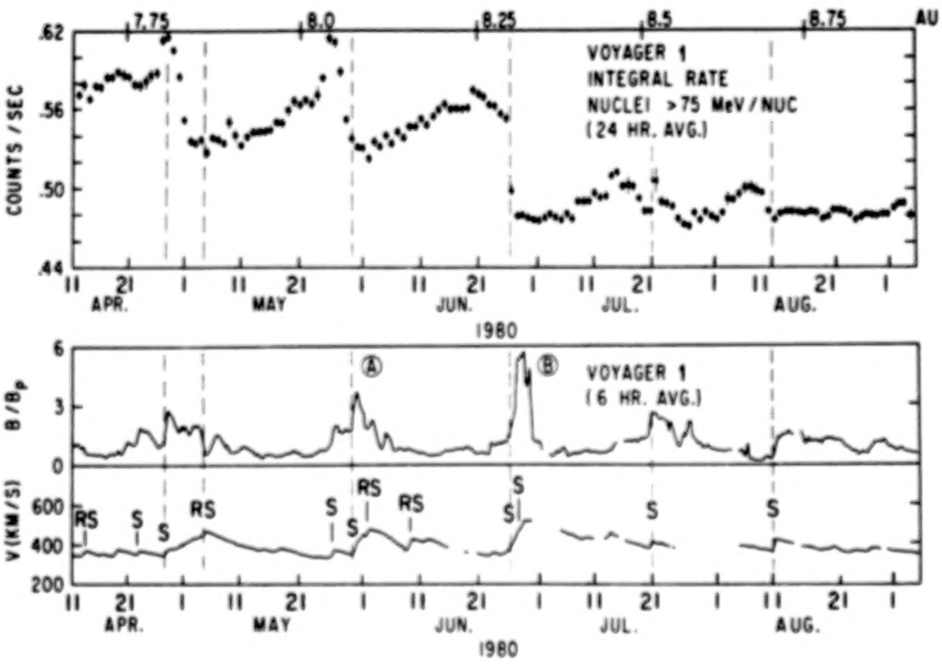


Figure 16. Comparison of cosmic ray observations by Voyager 1 with the magnetic field and solar wind velocity at a distance of approximately 10 AU. Compression regions at A and B are correlated with sudden decreases in the integral flux. [Burlaga et al., 1984]

studies have led to the current view that the transient high speed flows and disturbances originating at the surface of the Sun coalesce at large distances from the Sun to form essentially concentric shells of disturbed regions. These are the regions responsible for the modulation of the cosmic ray flux. This feature of the structure of the interplanetary medium on the large scale is illustrated in Figure 17.

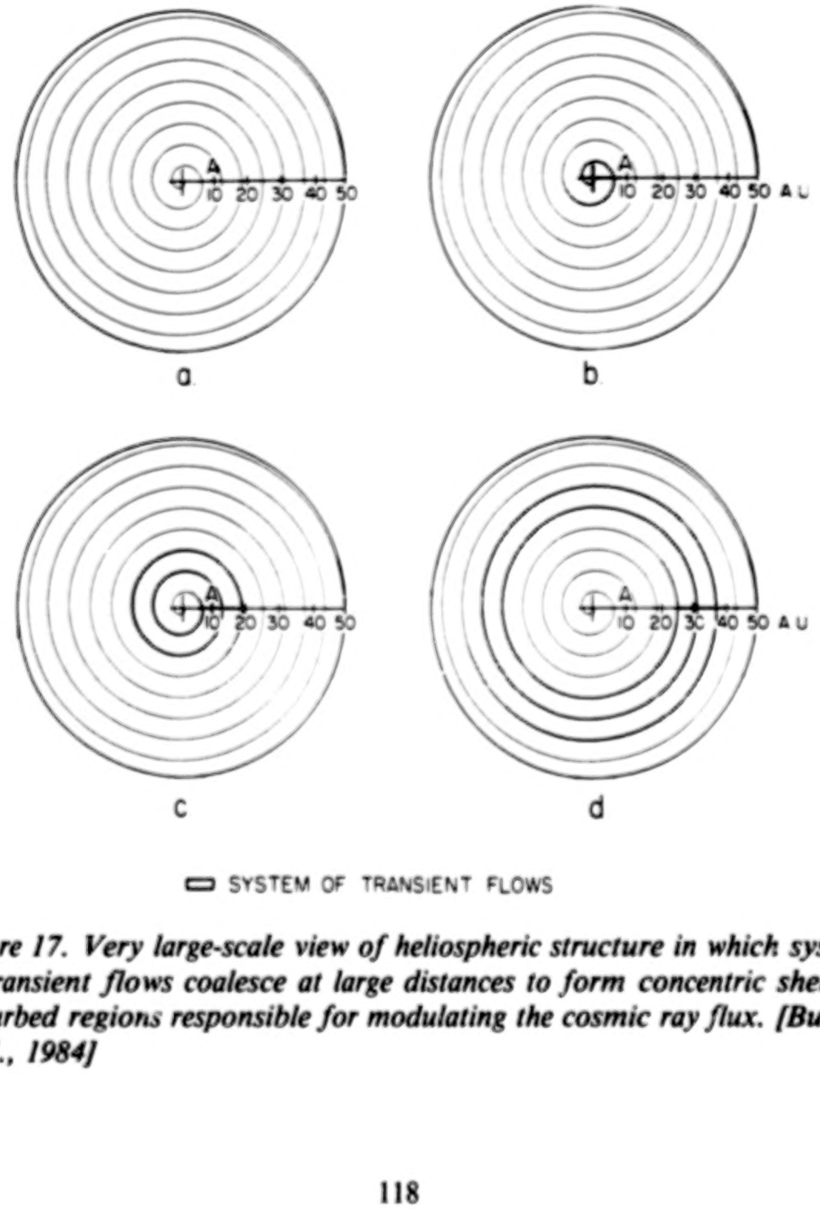


Figure 17. Very large-scale view of heliospheric structure in which systems of transient flows coalesce at large distances to form concentric shells of disturbed regions responsible for modulating the cosmic ray flux. [Burlaga et al., 1984]

5. SUMMARY

During the last two decades, spacecraft projects and individual experiments for which Frank McDonald has been a leader have contributed very significantly to our current understanding of the structure of interplanetary space and the correlation between solar and interplanetary disturbances. He had the foresight and ability to anticipate the unique value of small, simple, spin-stabilized spacecraft early in the NASA program.

For in situ observations of the interplanetary medium, this has proved critical. Studies on the IMP, HELIOS, and Pioneer spin-stabilized spacecraft and the larger attitude-stabilized Voyager spacecraft have provided unique data sets from which the modern view of the heliosphere has evolved. That concept is illustrated in Figure 18, in which the inner solar system is shown to be dominated by individual streams associated with specific source regions on the Sun. As these high-speed streams overtake the preexisting solar plasma, they coalesce and modify the characteristics so that at larger heliocentric distances, these disturbances appear as radially propagating concentric shells of compressed magnetic fields and enhanced fluctuations.

Frank McDonald has worked with a number of collaborators in his scientific investigations, both directly and indirectly. He has stimulated quality scientific investigations that have set a standard to which future investigations should aspire. These contributions will certainly stand the test of time as a remembrance of the efforts of FBM.

REFERENCES

Behannon, K. W. "Observations of the Interplanetary Magnetic Field Between 0.46 and 1 AU by the Mariner 10 Spacecraft." Ph.D. dissertation, Catholic University of America, 1976.

Behannon, K. W., Burlaga, L. F., and Hundhausen, A. J., 1983, *J. Geophys. Res.*, **88**, 7837.

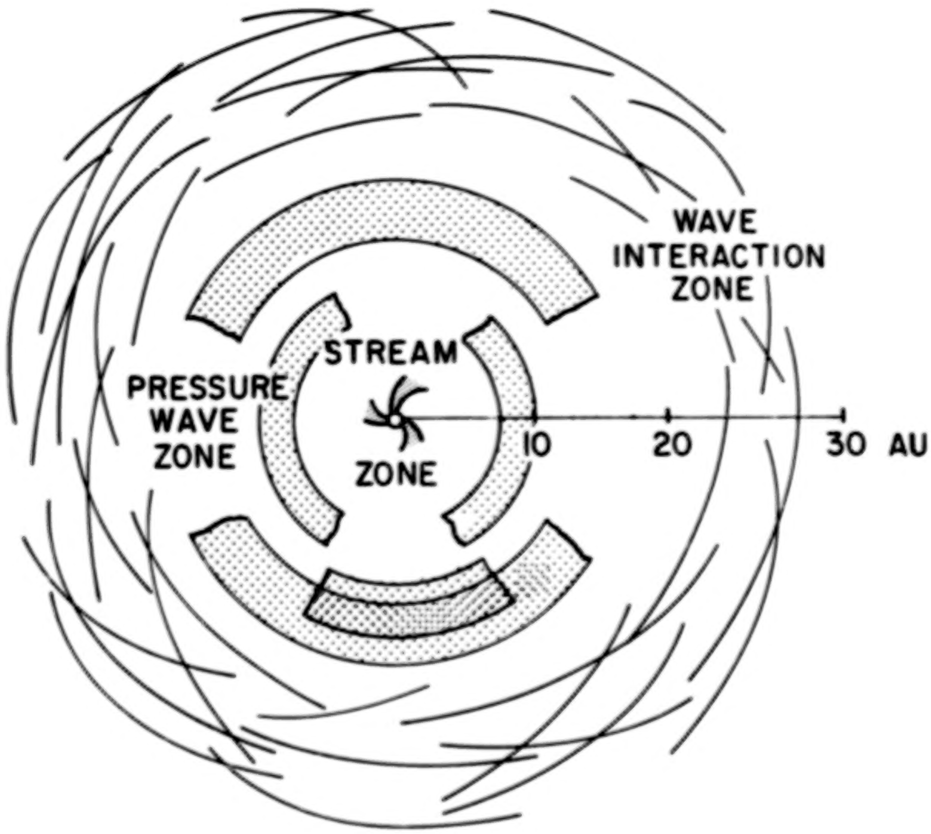


Figure 18. Diagram emphasizing the transition from an individual stream dominated zone within a few AU close to the Sun to a region of concentric shells where pressure waves dominate the structure, and, finally, the postulate that a wave interaction region exists beyond. [Burlaga, Schwenn, and Rosenbauer, 1983]

Burlaga, L. F., Hundhausen, A. J., and Zhao, J. P., 1981, *J. Geophys. Res.*, **86**, 8893.

Burlaga, L. F., 1983, *18th Internat. Cosmic Ray Conference Papers (Bangalore)*, **12**, 21.

Burlaga, L. F., 1984, *Space Sci. Rev.*, **39**, 255.

Burlaga, L. F., Schwenn, R., and Rosenbauer, J. H., 1983, *Geophys. Res. Lett.*, **10**, 413.

Burlaga, L. F., McDonald, F. B., Ness, N. F., Schwenn, R., Lazarus, A. J., and Mariani, F., 1984, *J. Geophys. Res.*, **89**, 6578.

Gosling, J. T., Hundhausen, A. J., and Bame, S. J., 1976, *J. Geophys. Res.*, **81**, 2111.

Ness, N. F., and Wilcox, J. M., 1965, *Science*, **148**, 1592.

Pizzo, V. J., 1978, *J. Geophys. Res.*, **83**, 5563.

Rosenberg, R. L., Coleman, P. J., and Ness, N. F., 1973, *J. Geophys. Res.*, **78**, 51–58.

Schulz, M., 1973, *Astrophys. Space Sci.*, **24**, 371.

PART II:

COSMIC RAY ASTROPHYSICS

N87
24254
UNCLAS

7

COSMIC RAYS IN THE HELIOSPHERE

William R. Webber
Department of Physics
University of New Hampshire
Durham, New Hampshire 03824

It is hard to believe—considering Frank's lofty position now and his long association with NASA—that in his early years he was extensively involved in the humble balloon flight program sponsored by the Office of Naval Research (ONR). This program led to many important early discoveries in cosmic rays during Frank's time at both Minnesota and Iowa. I can remember being in Iowa in the middle 1950's as a graduate student when Frank came down from Minnesota as a post doc. He was involved in building a "new" type of Cherenkov-scintilliation counter telescope to measure the spectra of protons and alpha particles. It sounded like interesting work so I began to work with him. In the first balloon expedition I remember, I recall loading all the equipment—experiment, gondola, and everything into Frank's old Stude to drive down to Texas. The Stude was one of the early models which looked about the same from the front and the back and really looked weird with all the equipment in it. I can remember stopping in some small Ozark town for gas—the woman looked at the car—then asked Frank what he did for a living. After a long pause he said, "I study cosmic rays." After an equally long pause she said, "I knew you were strange." Indeed, so this was really what cosmic rays were all about?

The period from 1955 to 1965 was one of balloon flights by the dozens to study cosmic rays from all kinds of places—from such exotic U.S. places as International Falls and Devils Lake, to foreign places like Guam. Those were the days when people launched their own payloads by hand, and there were many strange and bizarre tales to tell. One involving Frank that occurred in 1957 on Guam is worth telling. Frank's payload was being launched to measure

the proton and alpha particle intensity at >17 GV (still the best measurement available, by the way). Kinsey Anderson and I were holding the gondola aloft getting ready to run with it and Frank was behind us holding up the antenna. Cables, lines, etc. were everywhere. During the launch we had to run quite a ways to the side before we let the package go, and were too busy watching the balloon to notice Frank. After letting the package go we heard shouting, and all of a sudden Frank came sliding between us—feet up in the air, head sliding along the ground. He was tangled up in the load line and was about to be launched! Just as he left the ground his foot slipped loose and he fell back down—not much the worse for wear but completely ashen-faced. This incident has spawned two rumors—one not true—the other known only to Frank.

1. Frank lost his hair as a result of being dragged along the ground. (Not true—it wasn't there at the time of his arrival on Guam as verified by the arrival ceremony picture from the Guam Daily News, Figure 1.)

2. This incident caused Frank to leave ballooning and to look askance at it ever since. (As you know, Frank left Iowa to come to the newly formed GSFC soon after that.) Only Frank can answer this for sure, but I very much doubt it.

Nevertheless, this period in the late 1950's saw two important developments that owe much to Dr. McDonald. One was the development and refinement of the Cherenkov x scintillation telescope for the measurement of cosmic rays—a technique still widely used in various forms in both balloon-borne and spacecraft instrumentation. The other was the outgrowth of this telescope's ability to measure the spectrum of protons and alpha particles over a wide energy range and the extensive measurements that were made during periods of varying solar activity. In an important but largely forgotten paper, McDonald and Webber [1959], showed clearly for the first time that the energy dependences of the resulting cosmic ray intensity changes closely reproduced those to be expected if the cosmic ray changes were produced by a varying electric potential between the Sun and infinity. For a while several ideas involving solar electric potential models were considered as an explanation but a major step was made in the middle 60's when Gleeson and Axford [1968],

showed that Parker's diffusion convection model for solar modulation could be reformulated in what is known as the force field approximation in which the cosmic ray changes not only look like those produced by an electric potential, which the observations required, but that the particles actually incurred an equivalent amount of energy loss called adiabatic deceleration in their motion in the heliosphere through the outward flowing solar magnetic plasma. This truly marked the beginning of our modern understanding of solar modulation and the heliosphere.

SHEDDING OVERCOATS and enjoying Guam's warm weather on their arrival are seven of 13 scientists here to conduct operation "Skyhook", the Office of Naval Research cosmic ray expedition to begin soon in the Marianas. (L-R: Cdr. R. C. Cochran, Office of Naval Research Field Representative, Minneapolis; Lcdr. S. C. [illegible] ONR; Dr. R. E. Danielson, L. E. Peterson and Prof. Marcel Schein, University of Chicago; Robert C. Haymes, New York University; Dr. Frank B. McDonald, and Dr. Kinsey A. Anderson, University of Iowa. —NAVY photo

Figure 1. A few cosmic ray physicists arriving in Guam in 1957.

Frank's arrival at NASA coincided with the birth of the U.S. space program and studies of cosmic rays in earth orbit and beyond. Such spacecraft as the Orbiting Geophysical Observatory (OGO) and the Polar Orbiting Geophysical

Observatory (POGO) were legendary and one of Frank's important contributions was the development of the IMP series of spacecraft. These spacecraft contained McDonald's own brand of dE/dx × E telescope refined through the years so that it is now the backbone of small, compact spacecraft telescopes being used to measure cosmic rays. The IMP spacecraft are still operating today providing valuable data on cosmic rays and other interplanetary phenomena, and have truly lived up to their name "Interplanetary Monitoring Platform". The late 1960's and early 1970's was an interesting period in the development of our understanding of the heliosphere and cosmic rays. Mariner spacecraft measurements enroute to Mars had suggested a rather large interplanetary radial gradient (~ 50%/AU) of cosmic rays which coupled with other theoretical ideas suggested a scale size of the cosmic ray modulation region around the Sun (heliosphere) ~ 2-5 AU in radius. It was within this framework that the plans for the interplanetary probes called Pioneer and the later Voyager probes were spawned. Frank played an important role in the planning of these spacecraft and particularly the role that cosmic ray measurements would be given in the instrumentation of these spacecraft. An important year was 1973, just after the launch of Pioneer 10, as the cosmic ray community eagerly awaited the first results on the interplanetary gradient from three separate cosmic ray instruments onboard the Pioneer 10 spacecraft. First Van Allen reported measurements consistent with a zero gradient. What indeed was going on? First a 50%/AU gradient observed on Mariner and then one consistent with zero. Then a few months later, McDonald et al., 1974 (as well as Simpson at Chicago) reported well-defined gradients of only a few %/AU out to ~ 3 AU—e.g., Figure 2—and one immediately had a new perspective on a much larger heliosphere stretching to 10-20 AU at least! I can recall in the planning for the Voyager instrumentation that was going on in 1973 we still believed that this spacecraft would penetrate interstellar space at ~ 10-15 AU and tried to think of ways to determine that the spacecraft was indeed outside the heliosphere.

Of course now the two Pioneer and two Voyager spacecraft, launched in 1977 and all still operating, give us a truly interplanetary network of monitors, along with IMP, stretching from 1 to 35 AU, and in the case of Voyager 1 to nearly 30° out of the ecliptic plane (Figure 3). The scale of our modern heliosphere has now stretched to a radius of 50-100 AU (Figure 4). The general diffusion-convection models for cosmic ray modulation developed in the 60's are still

used—but now off-ecliptic effects, including drifts and dynamic effects not considered in the earlier spherically symmetric models, are considered important.

Any discussion of cosmic rays in the heliosphere must recognize several different particle populations, the study of which individually gives us somewhat different perspectives on cosmic ray motion in the heliosphere. These different populations include: (1) the Anomalous component, (2) low energy cosmic rays associated with co-rotating interaction regions (CIR's), (3) solar flare produced cosmic rays, and (4) galactic cosmic rays. We shall briefly describe these different types of cosmic ray particles and their role in the heliosphere. We begin with the Anomalous component, an area in which Frank has made some of his most important contributions. This story begins in 1973, after the launch of Pioneer 10, just as the measurements from the spacecraft were

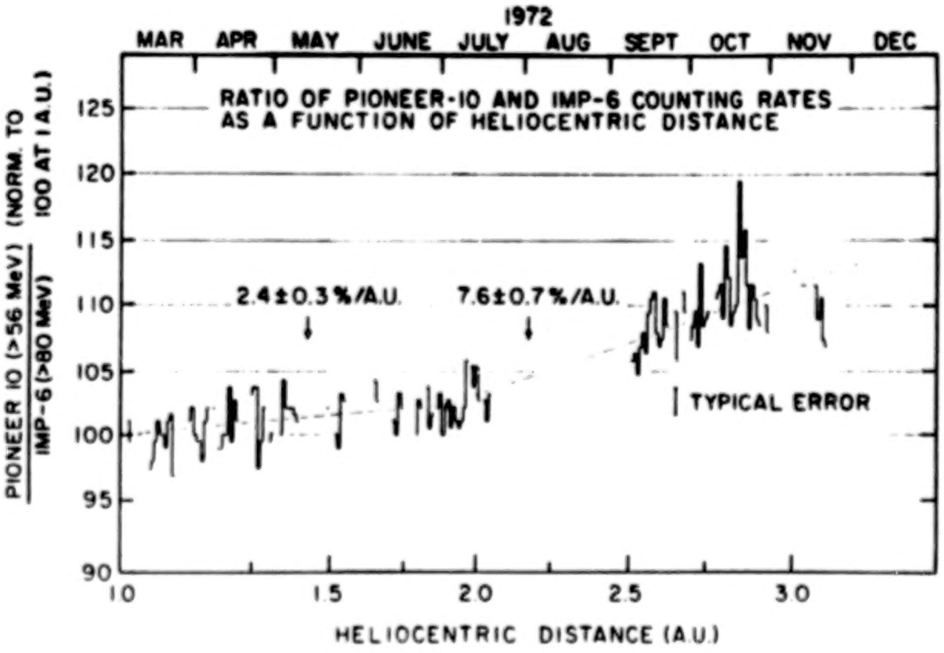


Figure 2. Early gradients measured by the Pioneer 10 space probe out to 3 AU as reported by Teegarden et al., 1973.

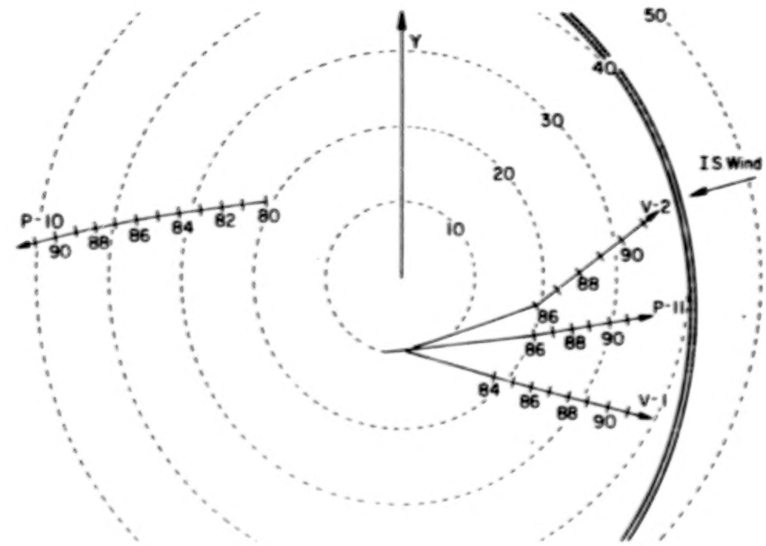


Figure 3. Locations of Pioneer and Voyager spacecraft within the heliosphere.

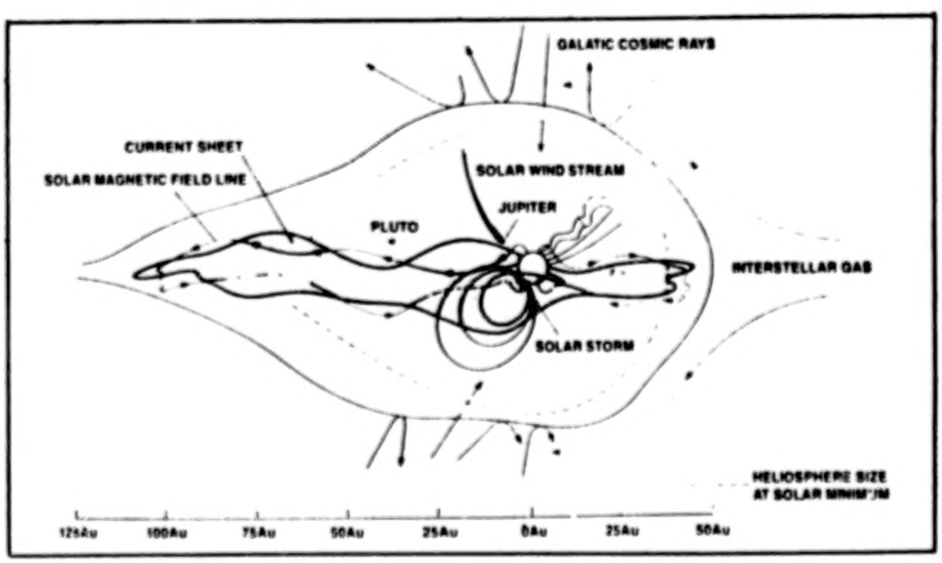


Figure 4. A modern artist's conception of the scale of the heliosphere.

beginning to define the radial gradient of cosmic rays as described above. Simpson and coworkers, Garcia-Munoz, Mason, and Simpson, 1973, had just reported an unusual flat helium spectrum seen first in 1972 (as opposed to one $\sim E^{1.0}$ expected from galactic cosmic ray modulation and seen earlier below $\sim$ 100 MeV). At the time of the Cosmic Ray Conference in Denver in August of 1973, McDonald brought graphs and tables showing an unusually large flux of nitrogen and oxygen but *not* carbon at energies $\lesssim$ 10 MeV/nuc. At the same time Hovestadt et al., 1973, reported an unusually large flux of oxygen at very low energies that did not appear to be related to solar flare activity. I personally believe it was Frank McDonald who convinced everyone at the conference that what they were seeing was indeed real and very strange and all of the different effects were related! [e.g., McDonald et al., 1974].

It soon became clear that all of the unusual fluxes of anomalous charges at low energies were elements with high first ionization potential. Typical spectra observed for the anomalous components in 1977 using Voyager spacecraft data near Earth are shown in Figures 5 and 6. The absence of anomalous C is clearly seen as is the presence of anomalous He which along with galactic He produces the flat He spectrum at low energies first reported by Simpson and coworkers. Various theories have been suggested with regard to the origin of these particles—including ones in which they are interstellar material accelerated in the heliosphere, in which case they should be singly ionized, to ones in which they come from a nearby galactic source. The most enduring of these theories, that they are singly ionized interstellar neutrals, was proposed by Fisk, Koslovsky, and Ramaty in 1974. There is now strong circumstantial evidence that these particles are indeed singly ionized and that they are accelerated somewhere in the heliosphere, but this view is not unanimously accepted—one of the nonbelievers being (I think) Dr. McDonald himself. Nevertheless, these particles have now been studied for more than an 11-year solar cycle. An example of the intensity variations seen at Pioneer 10 is shown in Figure 7. The data clearly show radial gradients and temporal variations that are remarkably different (and, in general, much larger) than galactic cosmic rays of the *same energy* coming to us from outside the solar system. Very significant changes in the spectrum of this component are also observed at the time of the solar magnetic field reversal in 1980. Data from the Voyager spacecraft show that the peak in the spectrum of anomalous O

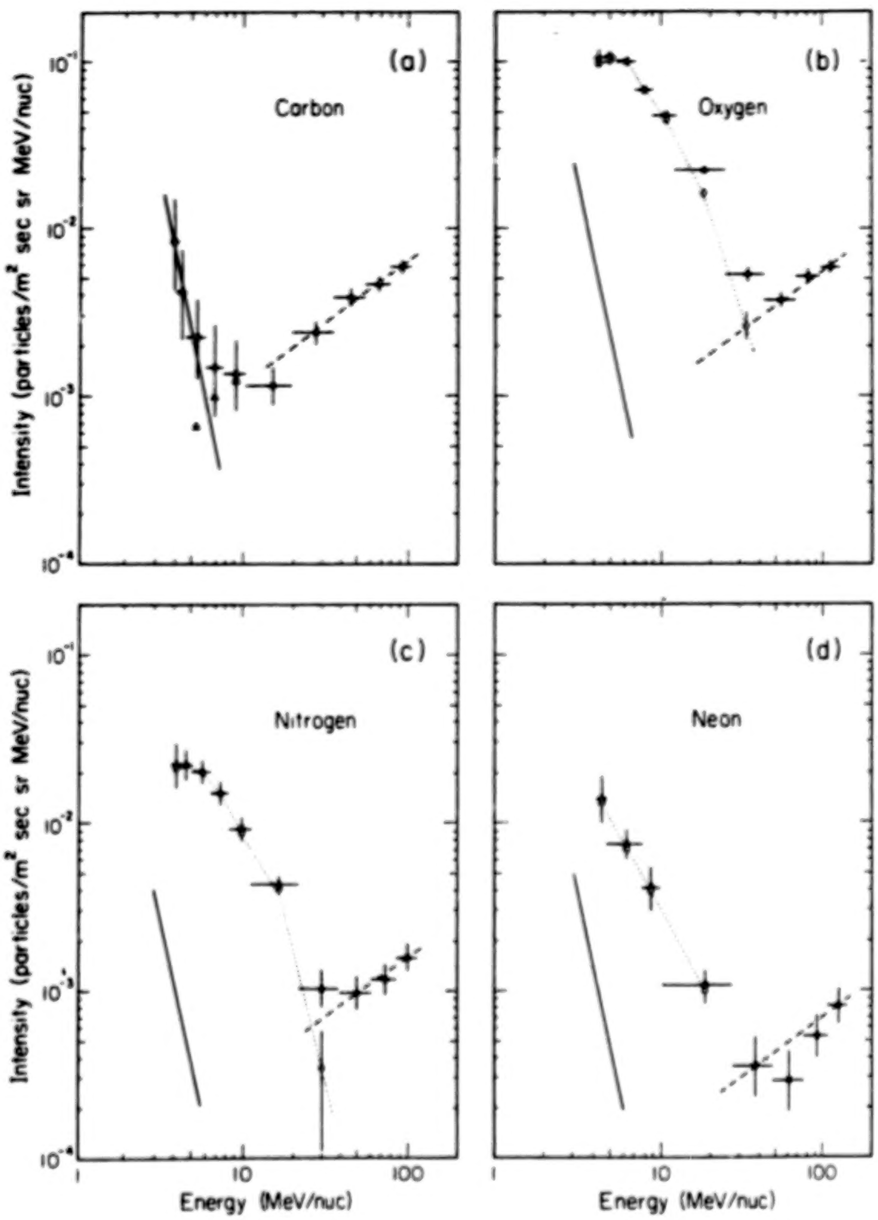


Figure 5. Examples of the spectra of C, N, O, and Ne at low energies showing the presence of anomalous N, O, and Ne.

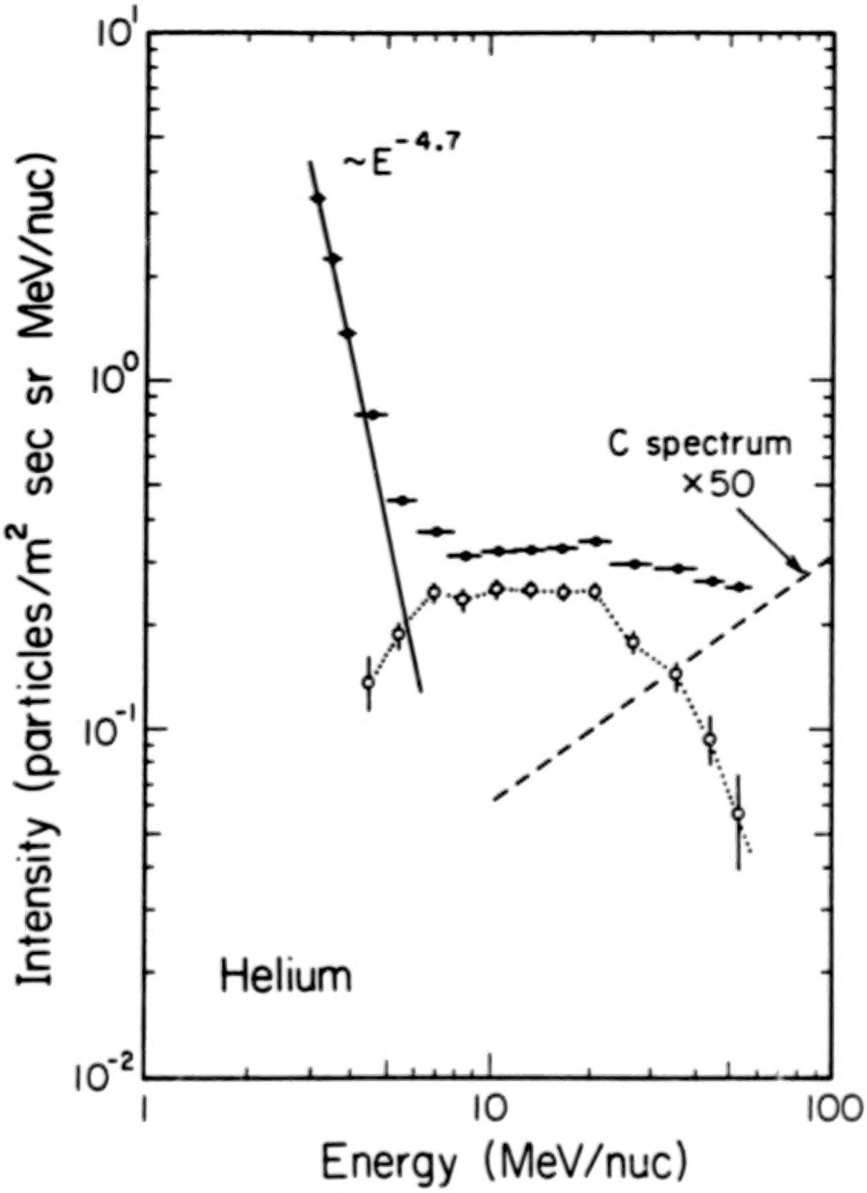


Figure 6. Helium spectrum at low energies showing separation into anomalous and galactic components between 10 and 100 MeV/nuc.

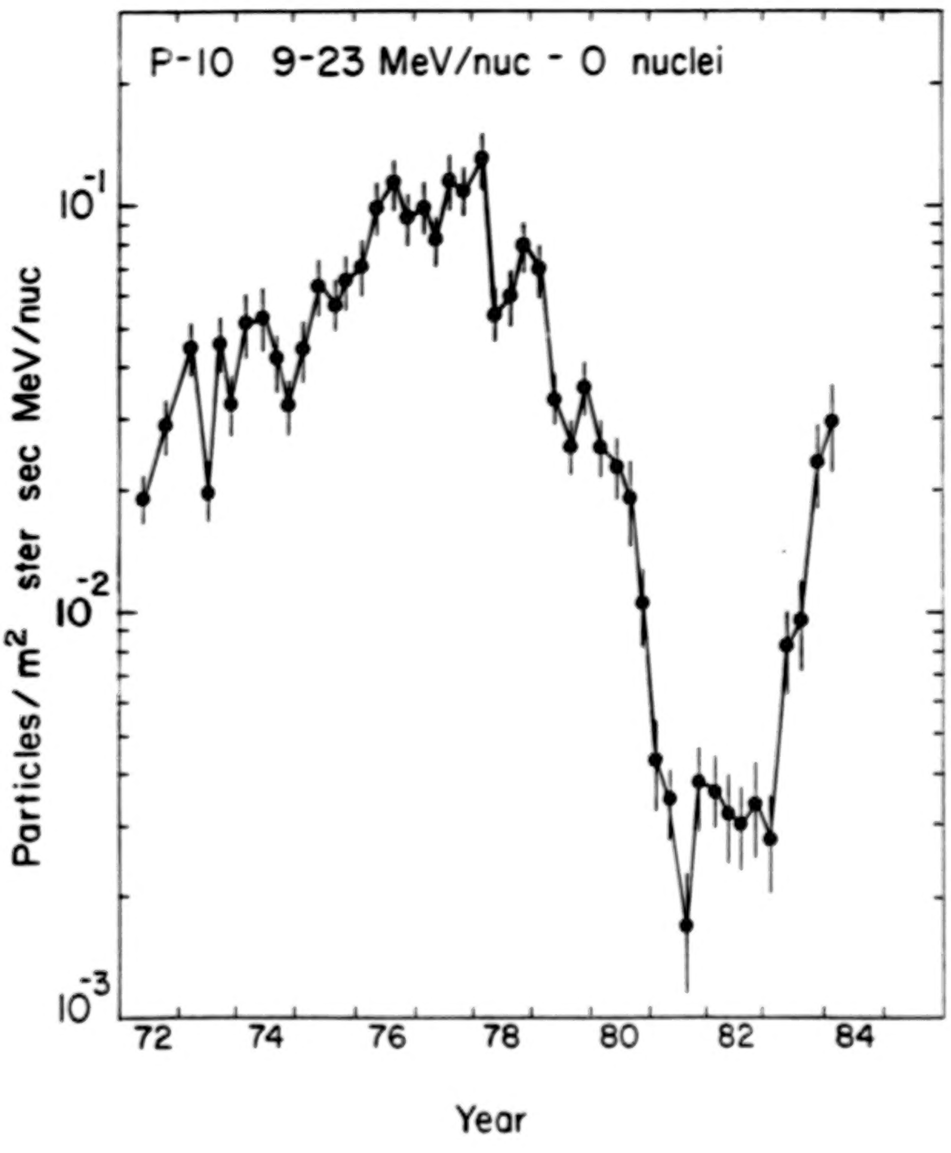


Figure 7. Temporal variations of anomalous O observed at Pioneer 10.

nuclei increases by a factor ~ 2 in energy after the field reversal in 1980 as illustrated in Figure 8. At the same time the radial gradient remained almost constant at ~ 10%/AU as illustrated in Figure 9.

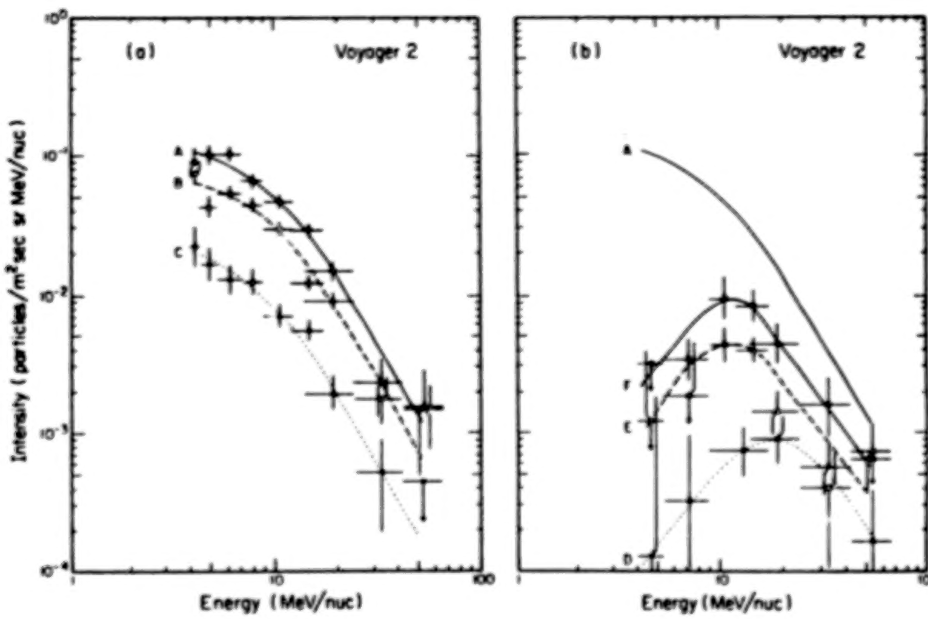


Figure 8. Voyager data showing the change in peak energy after the solar magnetic field reversal in 1980. Intervals A, B, and C are before the field reversal: intervals D, E, and F are after the field reversal. The large temporal variations of this component are clearly evident.

Solar system models for the acceleration of these particles have now moved to the boundary of the heliosphere. Models in which this acceleration occurs near the polar boundary of the heliosphere—accompanied by subsequent drifts to the ecliptic plane which change phase at the time of the 1980 field reversal—can account for at least some of the effects observed. However, the relatively constant gradient in time and space remains difficult to explain. It is clear that the study of this component is giving us a new and different perspective

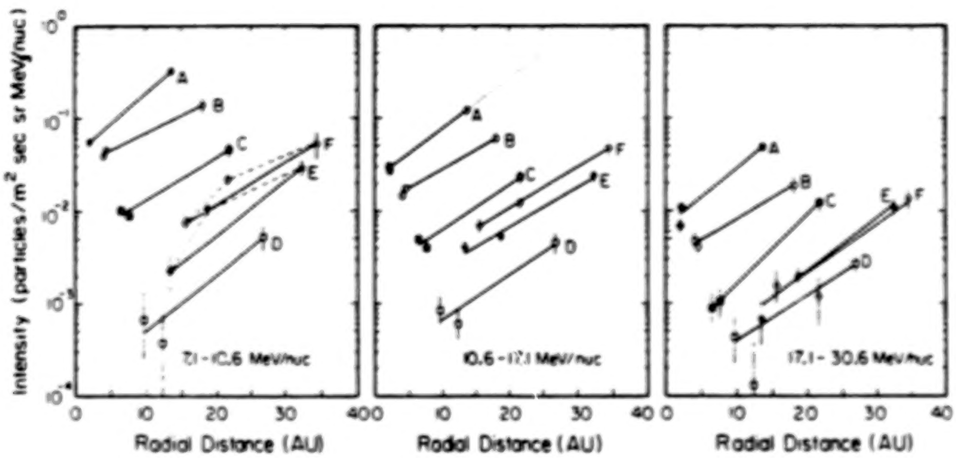


Figure 9. Intensity of various energy intervals of low energy O nuclei as observed at various radial distances and levels of solar modulation. A straight line with constant slope represents a constant gradient.

on the modulation of cosmic rays in the heliosphere, including possible effects caused by the boundary itself. At the same time, if these are directly accelerated interstellar particles as is believed by many, a firsthand example is provided of acceleration of interstellar material at heliospheric shocks, a process that must be very common in the galaxy.

A second cosmic ray population, a low energy component also accelerated within the heliosphere, (to which Dr. McDonald has also made important contributions) is that component associated with CIR's. This component was recognized for many years and originally identified with the arrival at Earth of an interplanetary blast wave/magnetic storm, following a large flare on the Sun which itself produced large fluxes of directly accelerated solar cosmic rays. These low energy particles were originally called energetic storm particles. Later spacecraft measurements convincingly related these particles to co-rotating interaction regions—their acceleration presumably occurring at the shocks bounding these regions and their presence being closely confined to these regions. Figure 10 is an example of these interplanetary particles

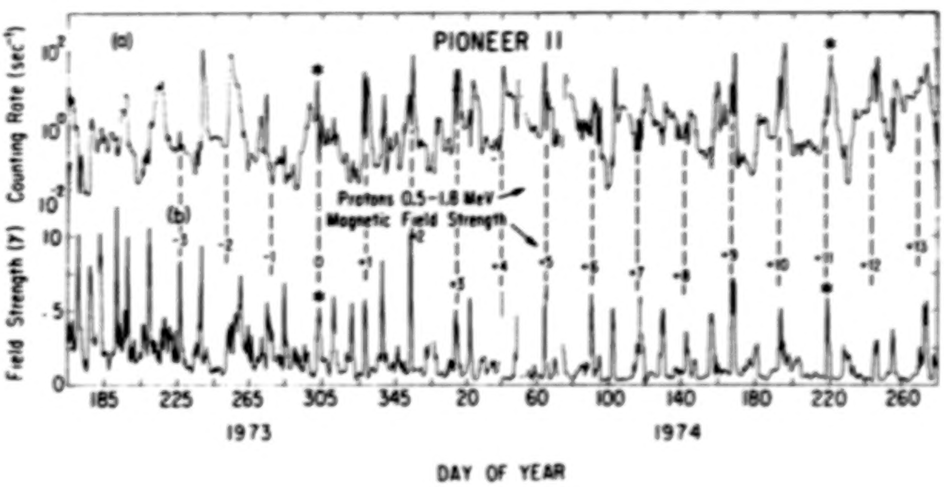


Figure 10. Co-rotating increases of low energy protons associated CIR's observed in 1973-1974.

(0.5-1.8 MeV protons) associated with a series of CIR's occurring in 1973-1974, many which recurred with a 27-day periodicity. Pioneer particle, magnetic field, and plasma measurements at different radii showed that these CIR's actually became stronger and more defined as one moved out from the Sun and faster moving shocks overtook slower ones, thus coalescing into fewer but stronger CIR's. In a landmark paper, Van Hollebeke et al., 1978, were able to follow individual events outward from the Sun using data from the Pioneers and IMP at three radial locations and to show that the cosmic ray intensity associated with individual CIR's reached a maximum at ~ 3 AU (Figure 11) as the strength of the shocks reached a maximum and then declined at larger distances out to ~ 10 AU as the shock strength slowly declined due to radial and azimuthal expansion. This direct connection between the cosmic ray intensity and the CIR strength was an important indicator that the cosmic rays were being accelerated locally in the interplanetary medium and *not* solar-accelerated particles trapped in the CIR's. Beyond 10 AU cosmic rays associated with CIR's or interplanetary shocks are observed less frequently but on occasion "giant" shocks are seen, coupled with enhanced fluxes of

cosmic rays up to 20-30 MeV, presumably accelerated locally. Examples of this type of event seen in April 1981 and in July 1982 at distances up to ~ 20 AU are shown in Figure 12.

Before leaving this topic we should mention another possibly related population of low energy cosmic rays—a very steep spectrum of cosmic ray protons, helium, and heavier nuclei that is present at quiet times. This component is illustrated in Figure 5 and particularly in Figure 6 for He nuclei. It seems to be present at all times—even magnetically quiet times and because

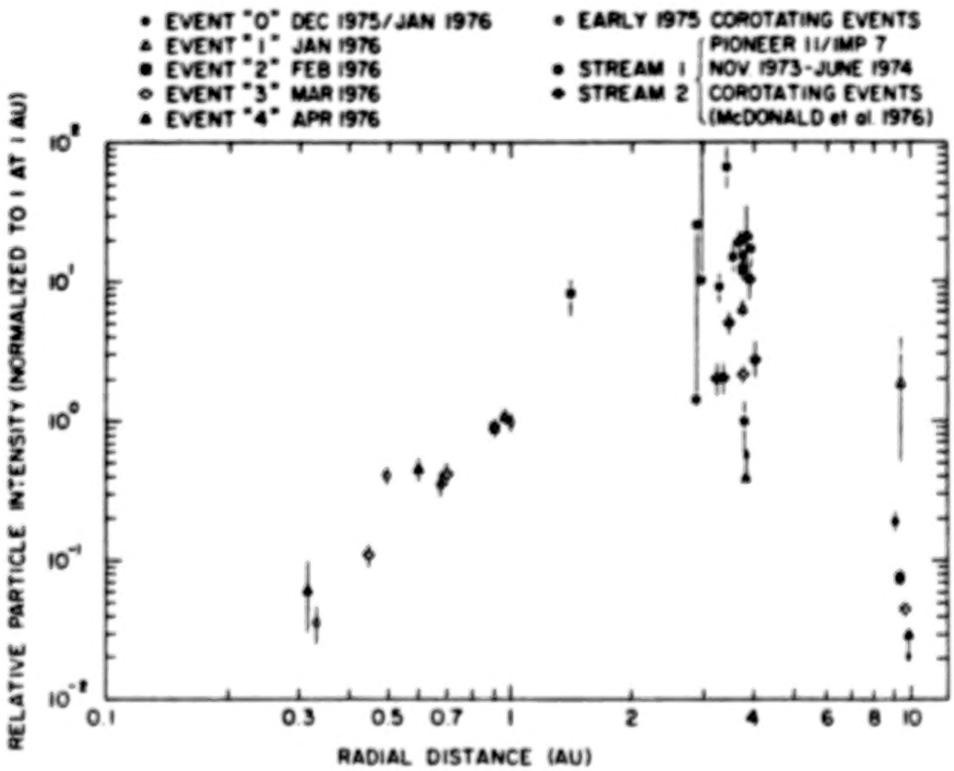


Figure 11. Relative intensity of low energy protons observed in co-rotating events at different radial distances showing a well-defined maximum at 3-4 AU.

ORIGINAL PAGE IS
OF POOR QUALITY

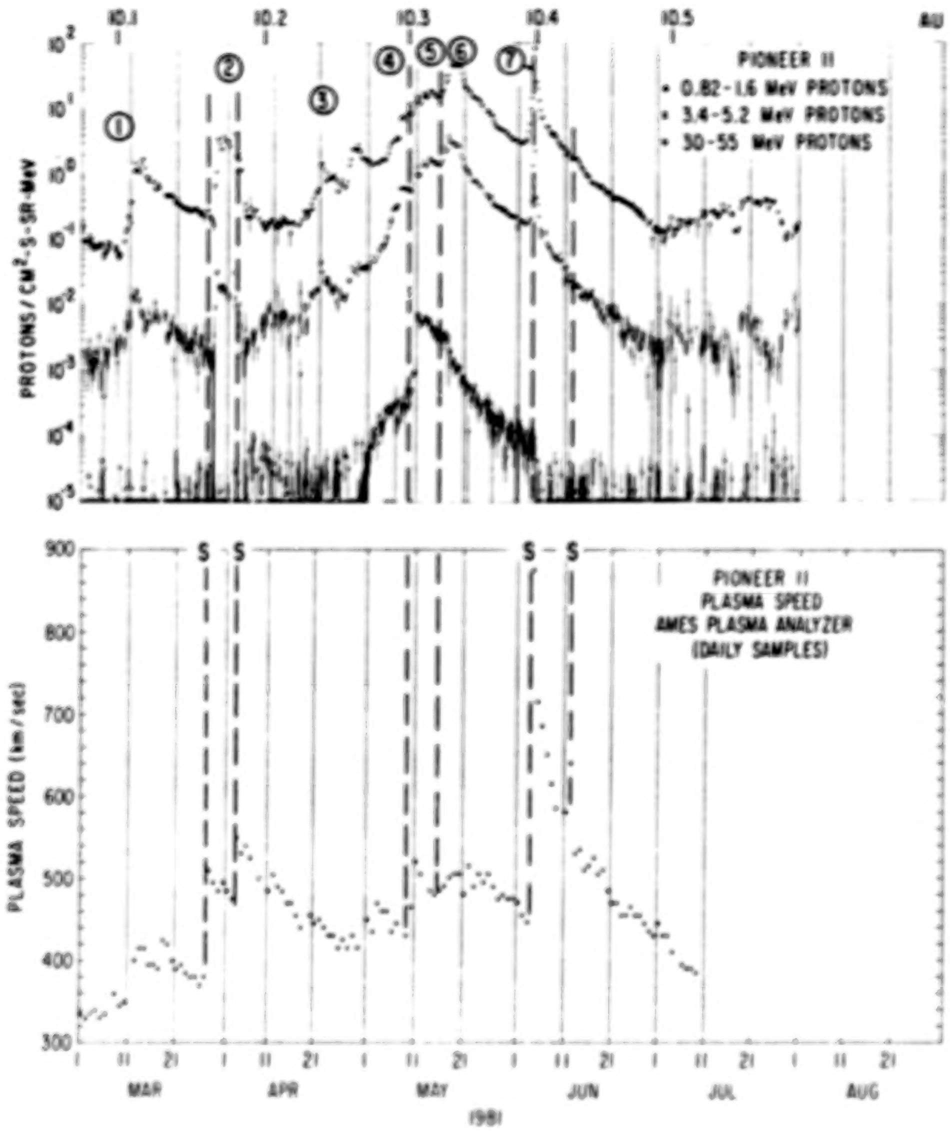


Figure 12. Low energy cosmic rays observed at Pioneer 11 (~ 10.3 AU) in April-June 1981. Examples of CIR-accelerated cosmic rays are labeled and are superimposed on solar flare cosmic rays from an event in April.

of its steep spectrum requires a considerable energy input to be accelerated and to be maintained against the adiabatic energy loss to be expected in the solar wind. It is observed at 30 AU, but no clearly established gradient has been defined. As yet it is not clear what its origin is—solar, interplanetary, or even at the boundary of the heliosphere. This forgotten population should certainly be a candidate for more study during the upcoming period of minimum solar activity.

A third, well-defined population of cosmic rays observed in the heliosphere is accelerated directly in solar flares—sometimes to energies > 100 MeV and on rare occasions to energies > 1 GeV when they are seen at sea level by neutron monitors. These particles have been studied in great detail since the great flare of 1956 with earthbased detectors. A major step forward in our understanding of the global properties of these particles has come from studies using spacecraft remote from Earth—particularly again the Pioneer and Voyager spacecraft. In terms of understanding energetic particle motion in the heliosphere, these particles have provided more local information—first on conditions between the Earth and the Sun and then later, after the launch of Pioneer 10, on conditions out to ~ 5 AU and beyond. Studies of the onset times, anisotropy, and intensity time profiles of these events have led to a picture of particle motion along magnetic field lines along with diffusion and energy loss that agrees with the picture obtained from the interplanetary gradient studies of galactic cosmic rays. At 1 AU these particles are clearly recognized by their intensity time profile, however, beyond ~ 5 AU the intensity of these solar particles diminishes greatly, and it becomes more and more difficult to distinguish them from the CIR-accelerated particles that appear as the flare-instigated shock propagates outward. An example of the complex behavior of low energy particles is shown in Figures 13 and 14. These figures show the intensity-time behavior of low energy cosmic rays during a six-month period in 1980 observed at the Voyager 1 and 2 spacecraft. These spacecraft are separated by ~ 1.2 AU in radius and have a small azimuthal separation. Temporal variations associated with CIR's are clearly evident with well-defined time delays between Voyager 2 and Voyager 1 associated with both radial and longitudinal propagation effects. A large solar flare increase is evident in the higher energies at both spacecraft in early August. This event is clearly related to a solar flare and to particles observed at Earth several days earlier. A good example of a large solar flare event occurring at Earth

ORIGINAL PAGE IS
OF POOR QUALITY

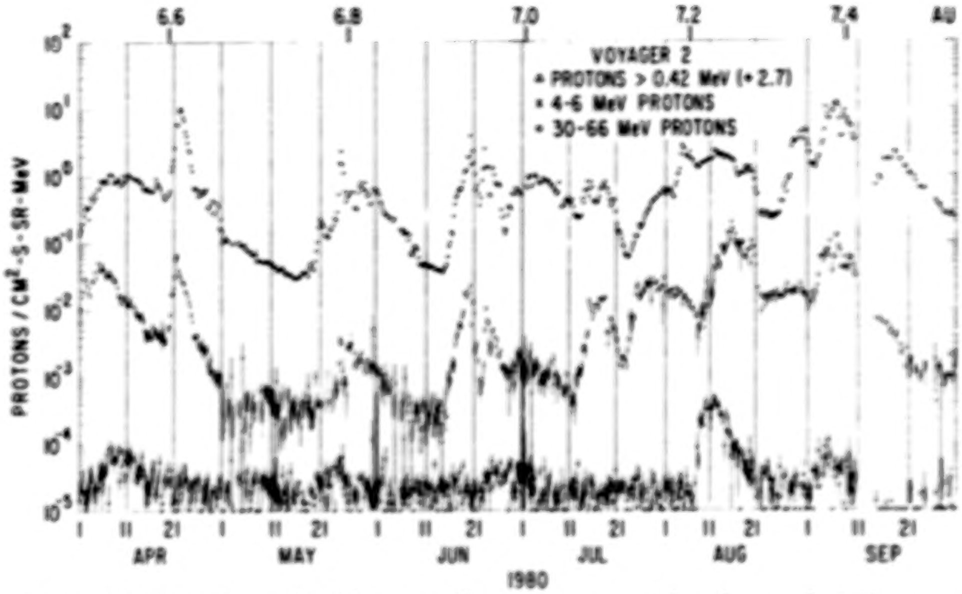


Figure 13. Time-intensity history of low energy particles observed at Voyager 2 in 1980.

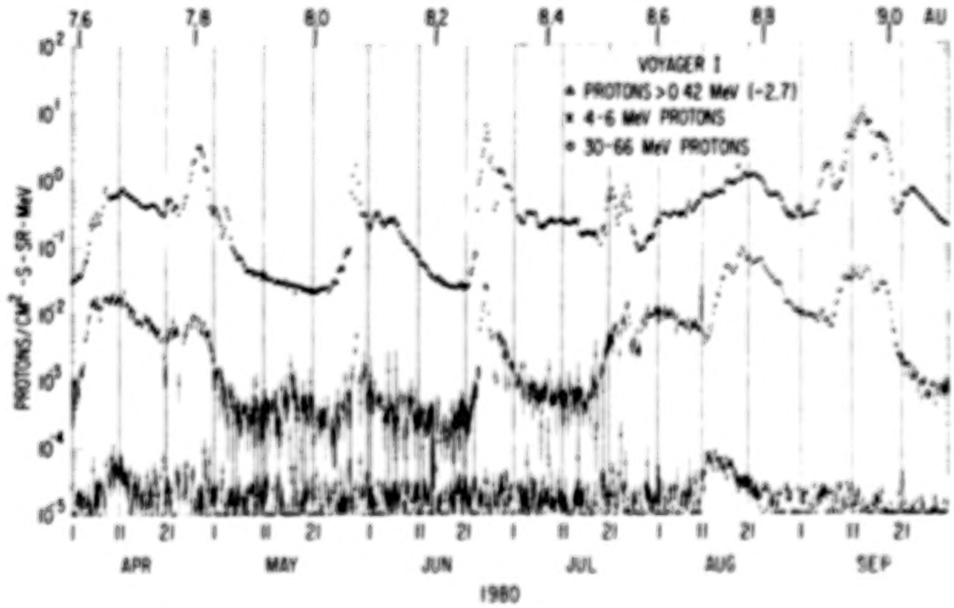


Figure 14. Same as Figure 13 except for Voyager 1.

in early June 1982 and seen at Pioneer 11 at 12.5 AU and Pioneer 10 at 28.1 AU is shown in Figure 15. At 28 AU the peak intensity is reduced by a factor $\sim 10^4$ and the event is spread over several months. The particles associated with the interplanetary shock clearly dominate the later phases of the event. These rare events that can be observed at large distances from the Sun provide some very sensitive tests of energetic particle propagation theory and need to be studied more carefully. As an illustration of how the parameters associated with these events scale with distance we show the time to maximum intensity and the total integrated intensity of particles as a function of distance in Figures 16a and 16b as derived from a study of over 10 of these events that could be identified at more than one radius. It is clear that beyond 30-40 AU it is unlikely that even the largest solar flare events will be observable above a few MeV, the intensities being decreased by a combination of energy loss and diffusion into a larger and larger volume. The early thought that solar-type stars could provide large quantities of energetic cosmic rays to interstellar space where they would be accelerated further to become galactic cosmic rays is thus unlikely—the original solar cosmic rays never make it out of the heliosphere for a variety of reasons, the principal one appearing to be adiabatic energy loss.

The fourth population of cosmic rays in the heliosphere is the galactic cosmic rays. These particles, incident on the heliosphere from outside, are energetically the most important population and are sensitively affected by the outward moving solar plasma and magnetic fields thereby producing the 11-year cycle of solar modulation. These particles have a long history of study at the Earth—here we shall dwell only on those studies remote from Earth that have helped to define the scale size and three-dimensional character of the solar modulation problem. The principal measurement that can be made in this regard is the interplanetary radial gradient—as a function of both energy and particle species if possible. An example of one of the types of measurement used to deduce the gradient is given in Figure 17 where the integral rates of $\gtrsim 60$ MeV protons measured by the telescopes on IMP, Voyager 1 and 2, and Pioneer 10 are shown as a function of time after the launch of Pioneer 10 in 1972. These rates are carefully normalized and show the large 11-year solar modulation effects beginning in 1978, as well as a growing separation of the individual rates, indicative of a radial gradient because of the progressive radial separation of the spacecraft. The gradient between Earth and Pioneer 10 is

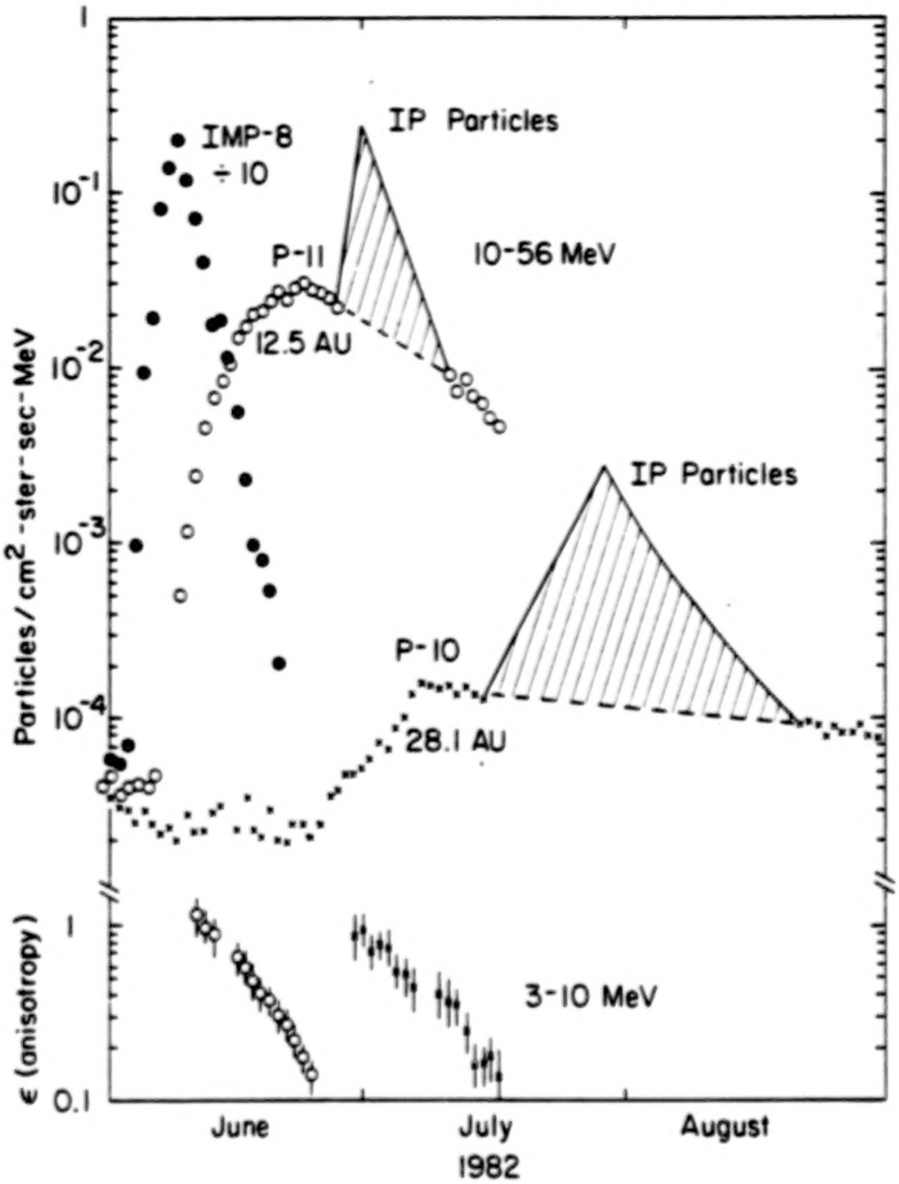


Figure 15. Intensity-time and anisotropy profiles for a very large solar flare event seen out to 28 AU from the Sun.

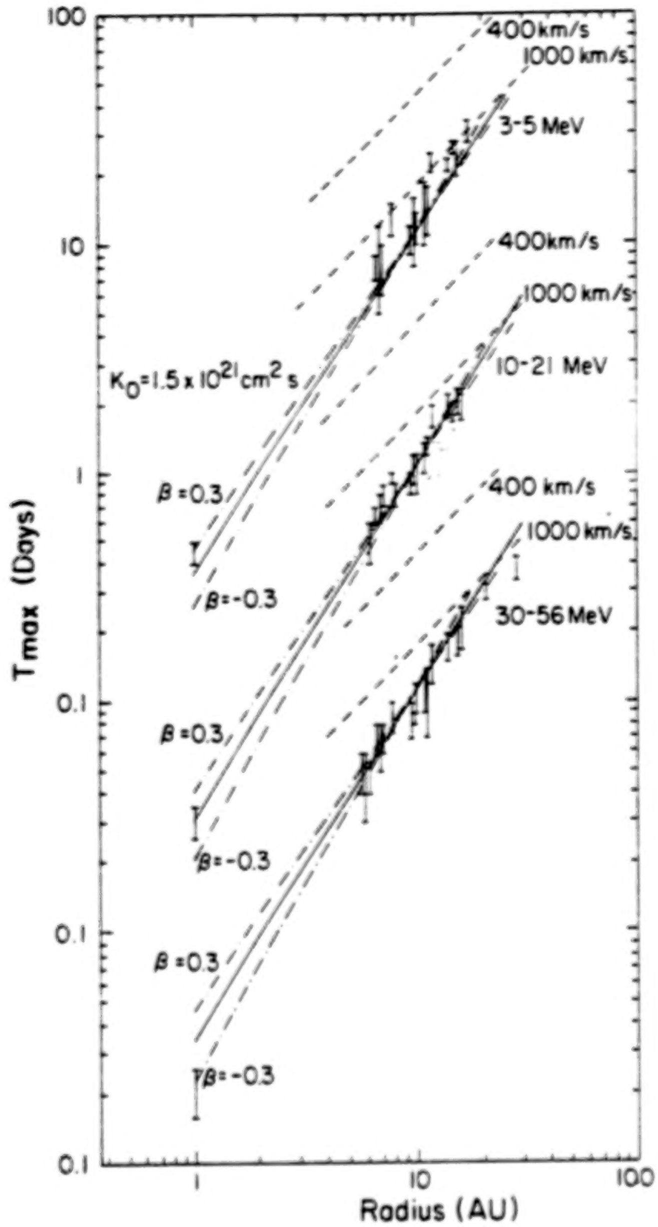


Figure 16a. Time to maximum intensity versus distance for solar flare events observed at more than one radius.

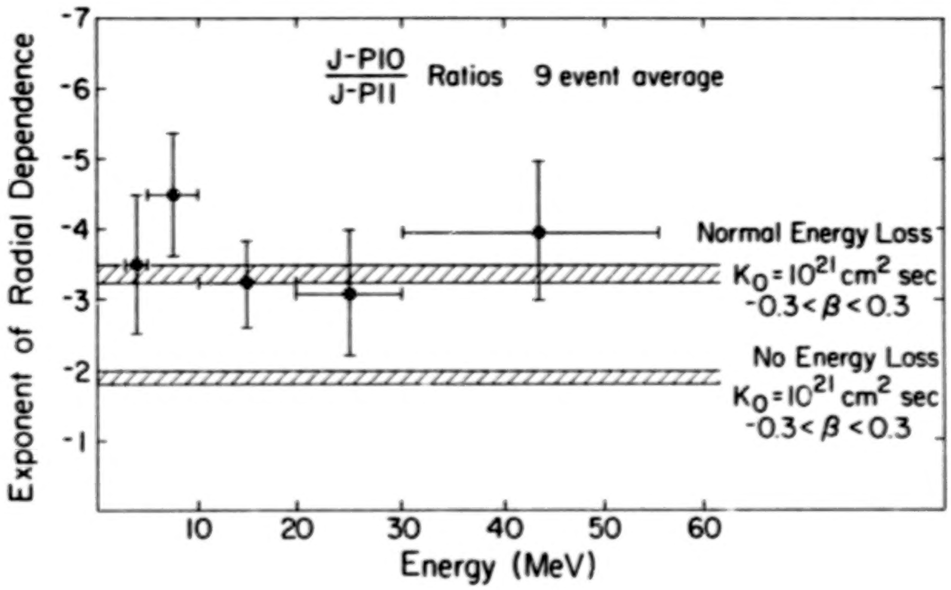


Figure 16b. Ratios of intensities as described by the radial dependence $[j_{10}(r_{10})/j_{11}(r_{11})]$ seen in the same solar flare at Pioneer 10 and Pioneer 11. Clear evidence for the effects of adiabatic energy loss is seen in this ration.

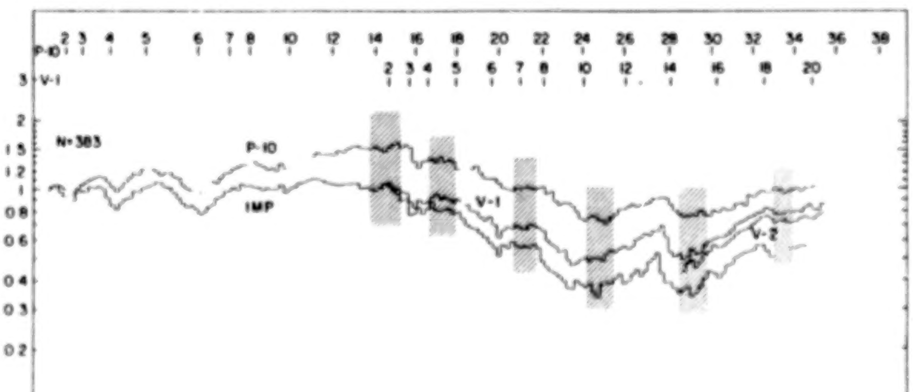


Figure 17. Intensities of > 60 MeV particles observed at different radii by instruments on IMP, Voyager, and Pioneer spacecraft.

illustrated in Figure 18. It remained constant for several years *and* as a function of radius to beyond 20 AU at a value ~ 2.8%/AU. Closer examination, using data from several spacecraft, has shown that there is probably a radial dependence of this gradient, (it is larger inside of 8 AU and appears to go through a minimum at ~ 10-12 AU) and that the general overall gradient has decreased considerably during the recovery phase of the solar cycle after 1982. Another way of illustrating the average behavior of this gradient is shown in Figure 19. The implications of the constant radial gradient, as well as the decreased (but still independent or r) gradient after 1982 are evident from this figure—most of the solar modulation must be occurring beyond 35 AU! Spacecraft are still well within the modulation region at 35 AU and it is unlikely that the boundary where the interstellar intensity is reached is closer than ~ 50 AU. In fact, the boundary location could vary with the level of modulation.

The rates of various energetic particles may be examined as a function of time and used to derive various differential energy gradients. An example showing ~ 120-250 MeV protons is shown in Figure 20. The radial gradient derived from this data (shown in Figure 21) shows a similar behavior with time as

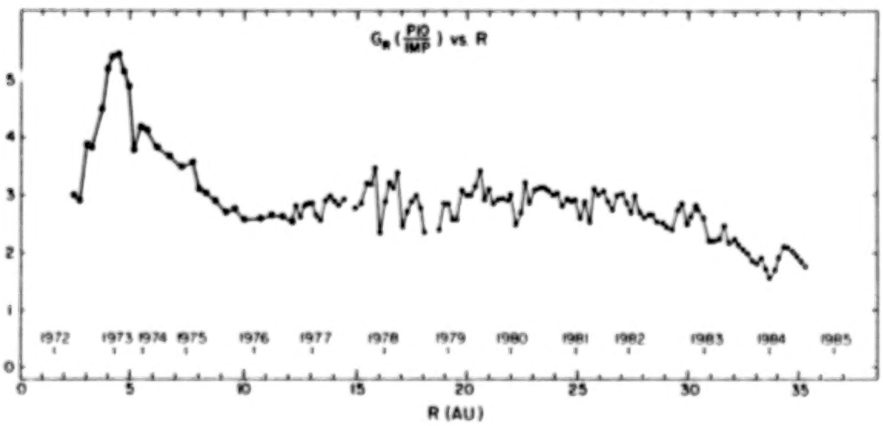


Figure 18. Radial gradient of > 60 MeV particles observed between Earth and Pioneer 10.

ORIGINAL PAGE IS
OF POOR QUALITY

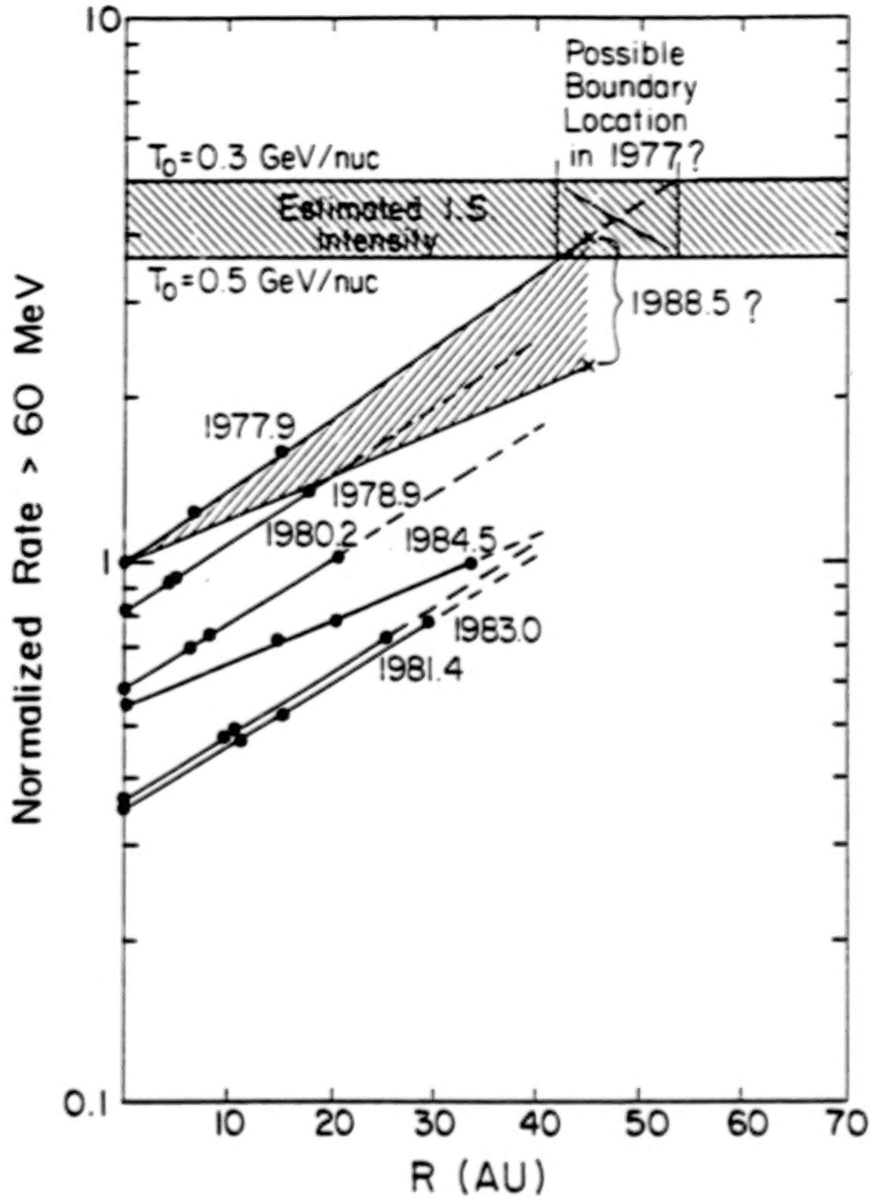


Figure 19. Normalized rate of > 60 MeV/particles as a function of radius at different epochs in time.

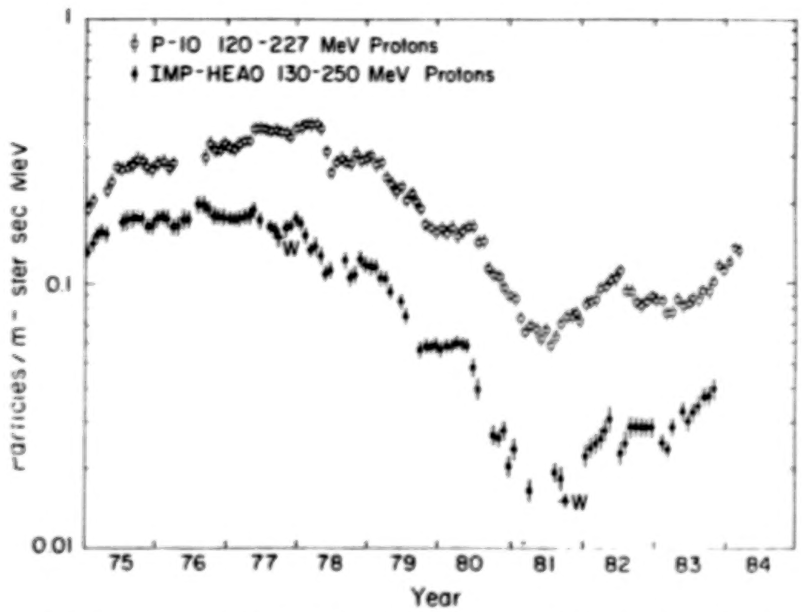


Figure 20. Intensity of 120-250 MeV protons observed by Pioneer 10 and IMP-HEAO.

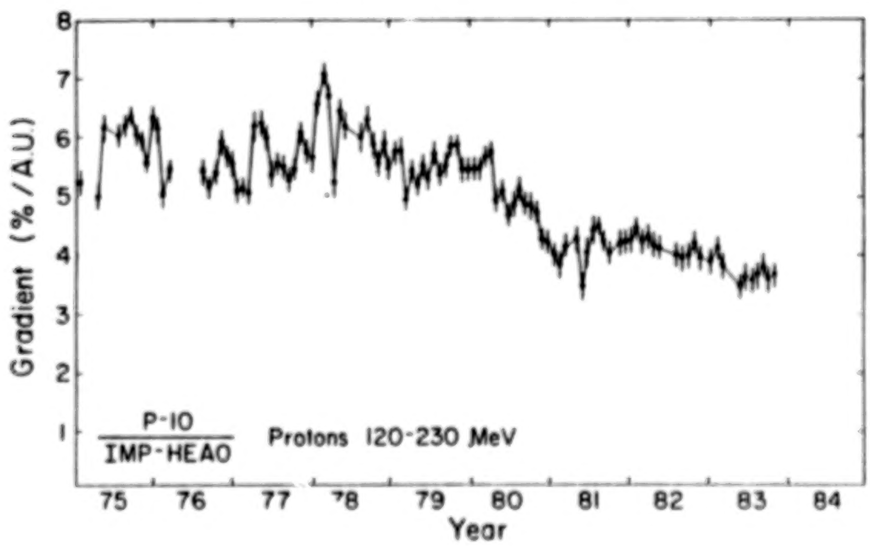


Figure 21. Radial gradient derived from the data presented in Figure 20.

the integral gradient, decreasing in amplitude after 1980. This kind of detailed analysis is unique to Frank McDonald. He is the only experimenter deriving the differential spectra and intensities of higher energy particles from spacecraft data and comparing intensities at different radii. This is proving to be very valuable. Examples of these detailed spectra that are a specialty of Dr. McDonald are shown in Figure 22. These comparative studies have shown that the gradients of all types and energies of particles have decreased dramatically after about 1982, not at the intensity minimum or at the solar magnetic field reversal in 1980-1981 but after the recovery phase of solar modulation was established in 1982. This is a new phenomenon and does not appear to have an immediate explanation in terms of the current modulation models.

One important question with regard to the dynamics of the 11-year modulation of galactic cosmic rays that has been answered using measurements from spacecraft at different radii concerns how this modulation cycle actually propagates: from the outer boundary in or from the Sun outward. McDonald was one of the first to show that this modulation cycle propagates outward from the Sun at approximately the solar wind velocity as is illustrated in Figure 23 [McDonald et al., 1981]. It has now been shown that the modulation cycle propagates outward at approximately the same velocity during the recovery part of the cycle as well. This suggests that part of the overall modulation is related to local phenomena and the modulation is a dynamic process, not fully considered in most earlier modulation models. One of these local phenomena of importance is the Forbush decrease, a large transient decrease in intensity associated with a blast wave or plasma disturbance emanating from a large flare or active region on the Sun.

One important aspect of trying to understand the overall solar cycle 11-year modulation of cosmic rays is the cumulative effect of many of these large transient decreases on the cosmic ray intensity. These decreases occur more frequently during the decreasing intensity phase of the cycle through minimum intensity. Models in which the 11-year variation is caused by a superposition of many Forbush decreases "piling up" in the heliosphere before they reach the boundary have been suggested. It has been possible to follow these Forbush decreases outward in the solar system to ~ 30 AU and beyond using Pioneer and Voyager data. An example of such a decrease occurring in July

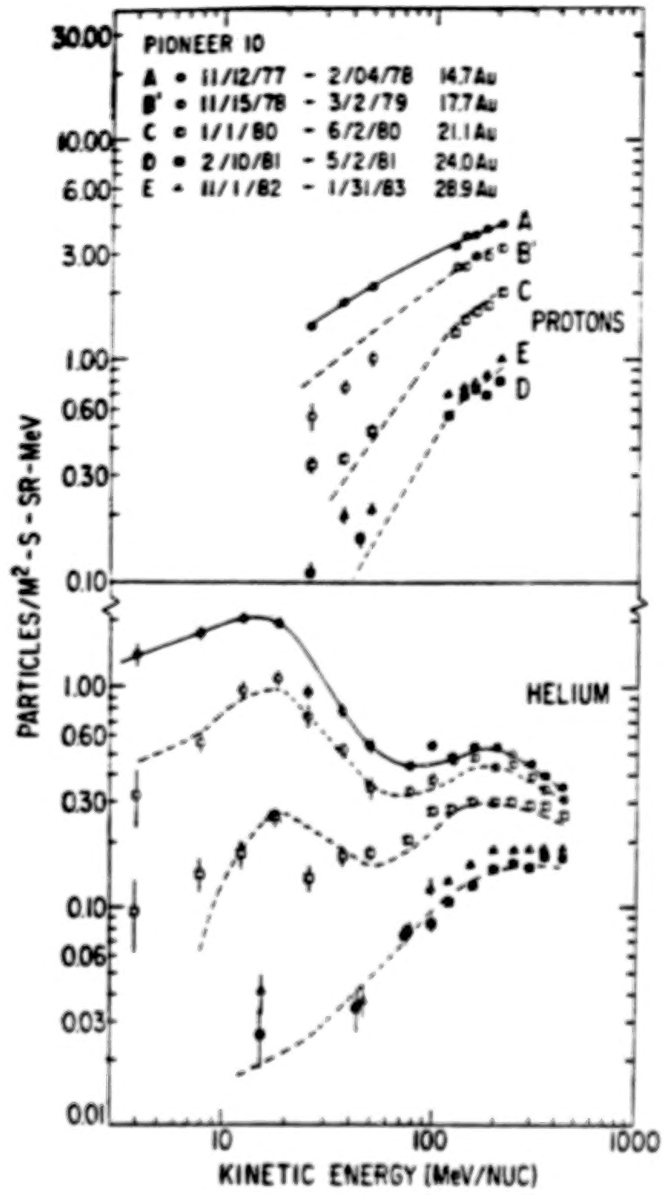


Figure 22. Energy spectra of protons and helium nuclei observed on Pioneer 10, showing temporal and radial variations.

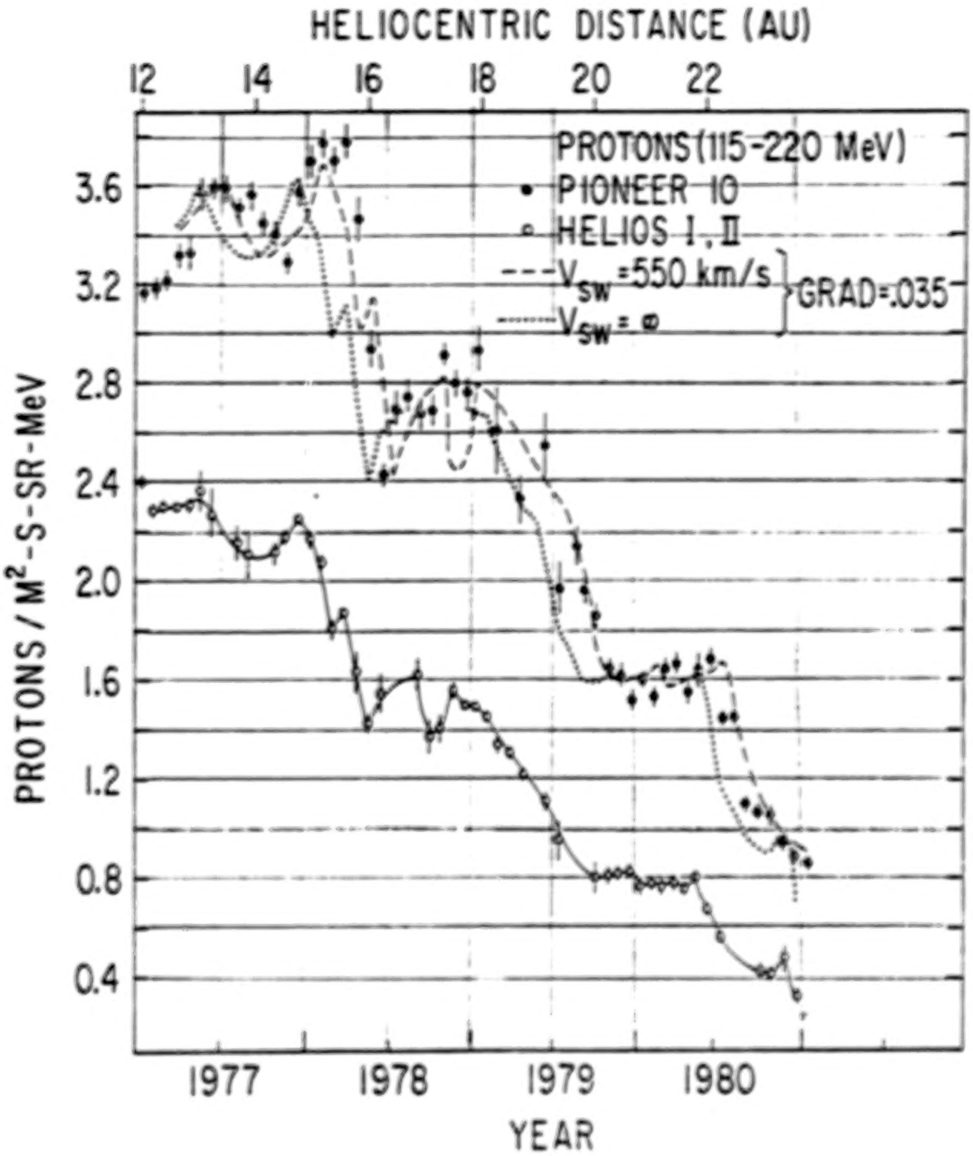


Figure 23. Intensity-time profiles of protons at Pioneer 10 and Earth, showing agreement when the profile at Earth is displaced by the travel time of the solar wind to reach Pioneer 10.

1982 and observed at several radii is shown in Figure 24. One important feature of this event is the longer recovery time at large distance. The characteristic recovery time varies from ~ 7 days at Earth to $\gtrsim$ 80 days at 21 AU. This feature is observed in ~ 20 Forbush decreases observed between 1978 and 1983 and, in fact, the recovery time is observed to be a function of radius, increasing to more than 100 days for events seen at large distances. If this kind of increase continues to 50 AU then recovery times $\gtrsim$ 1 year in duration might be expected and the characteristic recovery time scale from these decreases begins to look like that for the solar cycle 11-year variation. Perhaps the relationship between these two types of temporal variations is even closer than we think.

An important step in the understanding of the development of these transient decreases as they propagate outward has been achieved by Burlaga et al., 1985. They have built on their earlier idea that, as the shocks and magnetic field strength enhancements propagate outward, the faster streams overtake the slower ones producing a new kind of flow called a merged interaction region. These merged interaction regions dominate the transient variations at large distances and as a result the long term variation in intensity depends on the field strength in the interaction regions and the frequency of interaction regions, thus providing a direct connection to the 11-year variation.

The Pioneer and Voyager cosmic ray observations throughout the heliosphere are indeed giving us a new perspective on the three-dimensional character and scale size of the heliosphere. Most clearly they are emphasizing the role that transient variations in the outer heliosphere, and most likely the heliospheric boundary shock, play in the 11-year solar cycle modulation of cosmic rays. The next few years, as we pass through another sunspot minimum in 1988, are indeed crucial for interpreting and expanding upon these latest developments. If the Pioneers and Voyagers remain in good condition (and receive continued tracking), we will be able to sample the heliosphere to ~ 50 AU in both directions and to define more clearly the role of the outer heliosphere in all of these phenomena. One hopes that at Dr. McDonald's sixty-fifth birthday celebration in 1990, the quest for understanding the heliosphere and the role of cosmic rays in it which McDonald was instrumental in starting and pursuing with vigor, will be even closer to fruition.

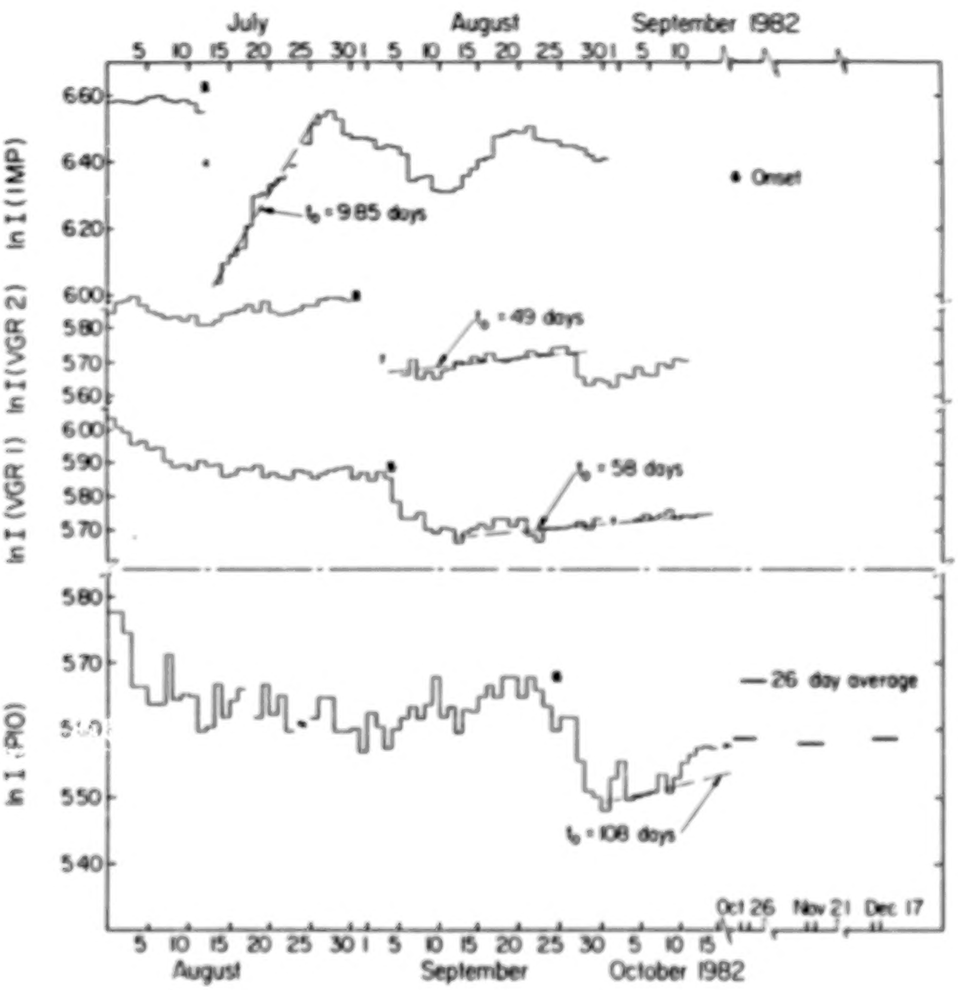


Figure 24. A large Forbush decrease observed by four spacecraft at different radial distances, clearly showing a longer recovery time at large distances.

REFERENCES (Selected Landmark References)

Burlaga, L. F., McDonald, F. B., Goldstein, M. L., and Lazarus, A. J., 1985, *J. Geophys. Rev.*, **90**, 12027.

Fisk, L. A., Koslovsky, B., and Ramaty, R. R., 1974, *Astrophys. J.*, **190**, 135.

Garcia-Munoz, M., Mason, G. M., and Simpson, J. A., 1973, *Astrophys. J.*, **182**, L81.

Gleeson, L. J., and Axford, W. I., 1968, *Astrophys. J.*, **154**, 1011.

Hovestadt, D., Vollmer, O., Gloeckler, G., and Fan, C. Y., 1973, *Phys. Rev. Letts.*, **31**, 650.

McDonald, F. B., and Webber, W. R., 1959, *Phys. Rev.*, **115**, 1974.

McDonald, F. B., Teegarden, B. J., Trainor, J. H., and Webber, W. R., 1974, *Astrophys. J.*, **187**, L105.

McDonald, F. B., Lal, N., Van Hollebeke, M. A. I., and Webber, W. R., 1981, *Astrophys. J.*, **249**, L71.

Teegarden, B. J., McDonald, F. B., Trainor, J. H., Roelof, E. C., and Webber, W. R., 1973, *Astrophys. J.*, **185**, L155.

Van Hollebeke, M. A. I., McDonald, F. B., Trainor, J. H., and von Rosenvinge, T. T., 1978, *J. Geophys. Res.*, **83**, 4723.

N87
24255
UNCLAS

N87-24255

8

ANTIPROTONS IN COSMIC RAYS

V. K. Balasubrahmanyan, J. F. Ormes, and R. E. Streitmatter
Laboratory for High Energy Astrophysics
Goddard Space Flight Center
Greenbelt, Maryland 20771

1. INTRODUCTION

Our experience with particle physics on the microscopic level at high energy accelerators has shown that particles are produced symmetrically with antiparticles. The extension of earlier nonrelativistic quantum mechanics to include relativistic effects led Dirac [1928] to predict the existence of positrons, subsequently discovered in cosmic ray experiments [Anderson, 1933]. High energy experiments with the collision of protons with other nuclei led to the discovery of antiprotons [Chamberlain et al., 1955]. These observations could be understood as conservation of baryon numbers and lepton numbers in nuclear interactions. These symmetry laws, as well as the phenomenon of annihilation of particle-antiparticle pairs when they interact, imply that if matter and antimatter exist in macroscopic quantities, they must be isolated from each other. The scale sizes of regions separating matter and antimatter are of fundamental importance in the study of cosmology [Stecker 1982, 1983; Steigman, 1976; and Zeldovich, 1965].

Recent experimental observations [Golden et al., 1979; Buffington, Schindler, and Pennypacker, 1981b; and Bogomolov et al., 1979] of antiproton fluxes larger than expected [Gaisser and Maurer, 1973] in the cosmic corpuscular radiation have stimulated the interest of physicists in several disciplines to consider their implications.

On astrophysical scales, the interfaces between regions of matter and antimatter might be revealed by the emission of gamma rays from annihilation processes.

However, other electromagnetic radiation from a source composed of antimatter would have identical characteristics to those from a source composed of ordinary matter. Consequently, X-ray, UV, optical, infrared, or radio observations are incapable of differentiating between sources composed of matter or antimatter. The cosmic radiation contains direct samples of matter from regions far beyond the solar system. Some fraction of these nuclei may be extragalactic in origin. If we were able to unambiguously identify samples of antinuclei ($Z \geq 2$) in the cosmic radiation (they cannot have been produced in collision processes), we would have an unambiguous signature of a large region of antimatter. Thus, cosmic ray composition measurements and gamma ray background observations have a very important bearing on the fundamental question: Is the universe symmetric in matter and antimatter?

2. EXPERIMENTAL RESULTS ON ANTIPROTONS

Table I, following Steigman [1976], lists experiments and techniques used to search for antiprotons and antihelium. Generally speaking, the techniques can be classified as either magnetic deflection or annihilation. The annihilation experiments can be further subdivided into those exploiting topology [Buffington, Schindler, and Pennypacker, 1981b] and those sensitive to the total energy release or calorimetry (emulsions).

The emulsion experiments have information on both energy and topology. In practice, however, the investigators [Apparao, 1967] look for an incoming slow proton-like track ($E < 200$ MeV) coming to the end of its range and causing a nuclear interaction of energy larger than the kinetic energy of the incoming particle.

Apparao [1967] looked for annihilation interactions in emulsion flown on balloons and from the absence of detection of annihilation interaction placed an upper limit for $\overline{P}/P$ at $< 9 \times 10^{-3}$ for rigidity < 0.6 GV/c.

Golden et al. [1979], using their balloon-borne superconducting magnet spectrometer, reported finding 46 antiproton candidates in the rigidity interval

TABLE 1
COSMIC RAY ANTIMATTER SEARCHES

	Rigidity (GV/c)	$\bar{N}/N$	Reference	Technique
	.5 to 1	$2.2 \pm 0.6 \times 10^{-4}$	Buffington, Schindler, and Pennypacker, 1981b	Annihilation Topology—Counter
	< 0.6	$<9 \times 10^{-4}$	Apparao, 1967	Annihilation—Emulsion
	< 1.3	$<3 \times 10^{-3}$	Aizu et al., 1961	Annihilation—Emulsion
Antiproton	3-6	$<1 \times 10^{-2}$	Bogomolov, Lubyanaya, and Romanov, 1971	Permanent Magnet
	5.6 to 12.5	$5.2 \pm 1.5 \times 10^{-4}$	Golden et al., 1979, 1984	Superconducting Magnet
	> 16	< 0.13	Durgaprasad and Kunte, 1971	Geomagnetic—Counters
	0.25 to 0.5	$< 2.2 \times 10^{-5}$	Buffington, Schindler, and Pennypacker, 1981b	Annihilation Topology—Counter
	1-10	$< 1 \times 10^{-3}$	Evenson, 1972	Permanent Magnet—Spark Chamber
	10-25	$< 8 \times 10^{-3}$	Evenson, 1972	Permanent Magnet—Spark Chamber
Antihelium	4-33	$< 5 \times 10^{-4}$	Smoot, Buffington, and Orth, 1975	Superconducting Magnets
	33-100	$< 2 \times 10^{-2}$	Smoot, Buffington, and Orth, 1975	Superconducting Magnets
	< 2.7	$< 7 \times 10^{-3}$	Aizu et al., 1961	Annihilation—Emulsion
	< 100	$< 7 \times 10^{-3}$	Damle et al., 1973	Permanent Magnet—Emulsion

5.6 to 12.5 GV/c. From these 46, 18 events were subtracted as due to atmospheric and instrumentation background. They interpret their data as resulting in a $\overline{P}/P = (5.2 \pm 1.5) \times 10^{-4}$. In a later publication Golden et al. [1984] have revised P/P to $(6.8 \pm 1.7) \times 10^{-4}$. In a joint experiment with the authors, Golden has plans to lower the threshold energy of his Cherenkov detector so that measurements with the balloon superconducting magnet can be made to lower energies.

Golden et al. [1984] have recently analyzed their balloon data to provide a differential spectrum of antiprotons in the few GeV range of energies. The poor statistics of the data make it difficult to see any clear pattern of correspondence of data with the various models. The data are more consistent with the shape of the $\overline{P}/P$ ratio expected from a secondary origin model than with a constant $\overline{P}/P$ ratio, but the flux is higher than expected by a factor of about 4.

Bogomolov et al. [1979] used a permanent magnet and spark chamber system to detect two events identified as antiprotons. Given the fact that the geometry factor of the system was only 1.1 cm^2, the derived flux is consistent with the results of Golden et al. [1979]. This result is limited by statistical uncertainties rather than by possible background effects.

The 1979 experiment of Buffington, Schindler, and Pennypacker found an unexpected result at odds with current theories of origin and propagation of cosmic rays. The energies of these observed antiprotons are below the production kinematic threshold and their flux is high, $\overline{P}/P = 2 \times 10^{-4}$. The results have been controversial. Detailed criticism of the experimental work has been made [see, for example, Stephens, 1981a]. Figure 1 shows the existing results on the measurement of the $\overline{P}/P$ ratio.

3. THEORETICAL INTERPRETATIONS

These antiproton observations, at the very least, force us to reexamine our current picture of the origin and propagation of cosmic rays, and they may imply evidence for more exotic processes.

We now survey the various theoretical ideas which have been suggested to interpret these unexpected experimental observations.

A. Secondary Production In Matter

The experimental data on the nuclear composition of cosmic rays has resulted in the development of several models for the propagation of cosmic rays in the interstellar medium [Cesarksy, 1980]. The crucial experimental observation of the (L/M) ratio of secondary nuclei (produced in nuclear collisions

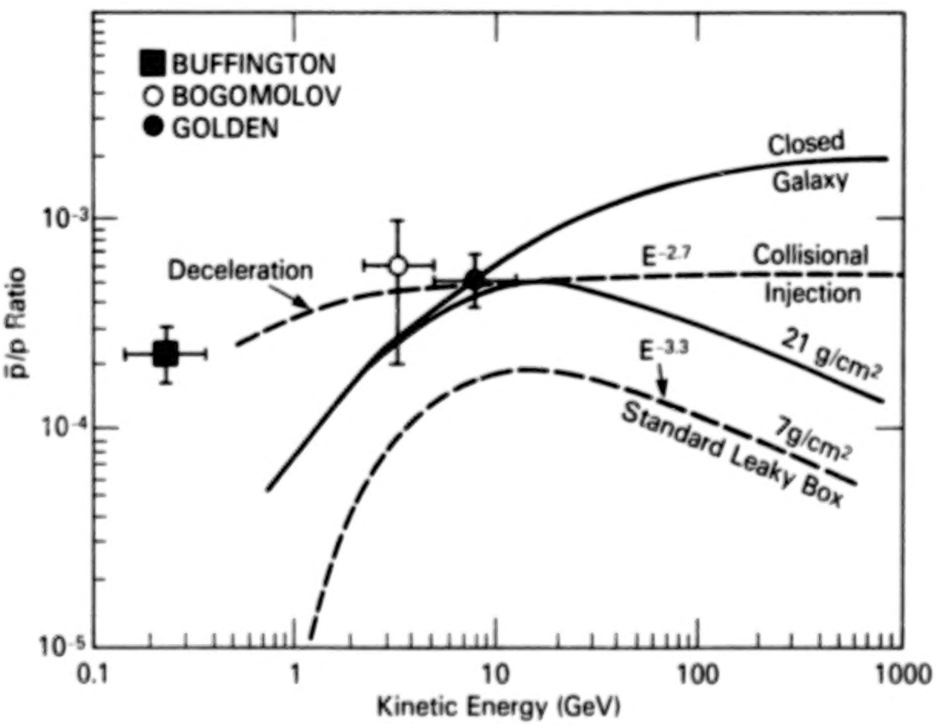


Figure 1. The data of Golden et al. [1979], Bogomolov et al. [1979], and Buffington, Schlindler, and Pennypacker [1981b] are compared with expectation from the class of models in which the antiprotons arise as secondaries of interactions.

of heavier nuclei with interstellar matter) to the primary nuclei has been the starting point of all the models. Antiprotons may be produced in collisions of protons with interstellar hydrogen. Given the observed proton flux, the $\overline{P}/P$ ratio can be calculated from known cross-sections and from the target thickness implied by studies of the heavier nuclei. An important feature of these studies is the variation of the target thickness as a function of energy. The behavior of the ratio at higher energies where no observations are available form the subject matter of the predictions from these models. In a separate paper [Cesarsky and Ormes, this volume] these models have been discussed, and for details the reader is referred to that paper.

In Figure 1 the observed $\overline{P}/P$ is compared to the predictions from the more prominent of these models. It is apparent that the $\overline{P}$'s are present in cosmic rays with a greater abundance than predicted [see curve labeled Standard Leaky Box]. The production of $\overline{P}$'s demands a much larger passage of matter than one would expect based on an analysis of the data from heavy nuclei. The curve labeled 21 g/cm^2 has been scaled up to fit the observations. This has motivated some workers [Cowsik and Gaisser, 1981; Mauger and Stephens, 1983; and Ginzburg and Ptuskin, 1981] to speculate on scenarios where more matter is traversed by cosmic rays in a certain phase of their acceleration. Essentially, these models suggest a separate source of protons (and possibly He nuclei), a source or sources surrounded by a large thick shroud ($\sim$ 50 g/cm^2). If heavy nuclei were accelerated in these sources, they would be broken up in this thick shroud of material and only protons and their secondaries will escape. (The mean free path for proton interaction is about 50 g/cm^2). Assuming these sources act like the sources of heavier cosmic ray nuclei, the predicted spectral exponent of these models for antiprotons is $E^{-3.3}$.

Recently Morfill, Meyer, and Lust [1985] have developed a model where shocks, produced by supernova remnants, interact with nearby clouds. The enhanced cosmic ray abundances accelerated in the shock produce secondaries when the shock interacts with the cloud. If clouds fill 8% of the interstellar medium and hot, low density gas makes up the remainder of the medium, they claim an agreement between the calculated secondary to primary ratios with observations. The energy dependence of the secondary to primary ratio comes from the energy dependent escape from the waves near the shock. This

model predicts that the $\overline{P}/P$ ratio will fall with increasing energy up to some energy (perhaps about 100 GeV) at which the ratio will flatten to a component due to interactions with the averaged interstellar material. This paper did not present quantitative predictions.

Another class of models separates the origin of protons from the heavier nuclei. These are generally models in which the population of protons (and perhaps helium) is old and has traversed more matter than the heavy nuclei which are presumably younger and produced in a nearby source. These are sometimes known as closed galaxy models [Peters and Westergaard, 1977] (closed galaxy curve). Protheroe [1981] has shown that the antiproton flux in the energy range above 1 GeV is consistent with the predictions of the closed galaxy model, but the low energy data are inconsistent with this model. Stephens [1981] has postulated a three-tier model which is a combination of the closed galaxy and the leaky box models. About one-half the protons are young and reach us promptly from the spiral arms whereas the other half is trapped in the outer galaxy. Being trapped, they traverse a lot of matter and so produce a larger amount of secondary antiprotons. Stephens' model is capable of matching the observations of both P and e^+. These models generally produce spectra which are not power laws as they admix components with different exponents.

The observed antiproton data demand not only large matter traversal but also a mechanism of energy degradation from the GeV range to a few hundred MeV, as pointed out by Buffington and Schindler [1981a], Eichler [1982], Ginzburg and Ptuskin [1981], and Mauger and Stephens [1983].

There is a class of models in which antiprotons are produced in collisions, and then injected into an accelerator along with protons. Those models produce the same asymptotic spectra for protons and antiprotons.

Ginzburg and Ptuskin [1981] have considered production of antiprotons in young supernova envelopes where cosmic ray protons pass through appreciable amounts of matter. The antiprotons would undergo adiabatic energy losses in turbulent regions in the envelope and their spectrum is weighted towards low energy. These regions, being surrounded by large amounts of matter, would not let heavy nuclei leave the sources, as they would break up by nuclear and photonuclear reactions at this active stage. Nuclei could be accelerated at a

later stage from these remnants or as Cowsik and Gaisser [1981] postulate, the sources of heavy nuclei and protons and He nuclei could be different.

Moraal and Axford [1983] and Mauger and Stephens [1983] produced the antiprotons in the early dense phase of a supernova explosion. The details differ, but these models feature the injection of antiprotons at low energy as secondaries and then accelerate them along with protons so the spectra should be the same ($E^{-2.7}$) at high energy. (See curve labeled Collisional Injection.)

Tan and Ng [1983] attempt to interpret the observed antiprotons as arising from the interactions of protons in the dense molecular H_2 cloud regions concentrated in a ring of radius 5 Kpc around the galactic center. The gamma ray data and molecular H_2 density distribution derived from molecular CO distribution lend credence to a nonuniform density distribution in the galactic disc. Tan and Ng claim that all the available data on antiprotons, including the low energy data of Buffington, Schindler, and Pennypacker [1981b], can be explained by their model. The antiprotons are secondaries produced in the 5 Kpc molecular ring. Subsequent adiabatic deceleration due to the expansion of the ring decreases their energies below the kinematic threshold. Secondaries of heavier nuclei are produced on scales comparable to the disc and are not affected by these special effects. If matter concentration is important for the production of antiprotons, it should have important contributions to secondaries and of Fe and ultraheavies. Unless there are special sources of antiprotons as Cowsik and Gaisser [1981] speculate, it is hard to discount effects on heavier nuclei. What happens to He nuclei, for example, in this model? Does one expect a large $^3He/^4He$ ratio at low energies? Details are not available from the work of Tan and Ng.

Lagage and Cesarsky [1985] have examined the general problem of explaining antiproton fluxes by production in thick sources. They conclude that while such sources may contribute as little as 25 percent of cosmic rays, minimum source grammages of about 30 g cm^{-2} are needed to avoid production of light secondary nuclei in excess of observation. They calculate that gamma rays from these sources would then be expected to contribute somewhat more than half of the observed gamma ray flux above 100 MeV. This is barely

tolerable as one can already account for at least half of the gamma ray flux by diffuse emission from cosmic rays interacting with ambient interstellar medium [Lebrun et al., 1983]. The problem of overproduction of gamma rays in thick sources is exacerbated if cosmic rays are adiabatically decelerated in the sources, increasing the source luminosity.

Only models with collisional injection are able in a natural way to explain the abundance of antiprotons at 200 MeV. Other models must add substantial deceleration in some manner or other to produce antiprotons an order of magnitude below the kinematic threshold, even including solar modulation effects.

Dermer and Ramaty [1985] investigated the possibility that antiprotons are produced in (p-p) collisions in relativistic plasmas. In their model, both projectile and target protons are in motion, and the antiproton production kinematic threshold is lower in the frame of the plasma. The spectrum in this case would extend to much lower energies compared to the production of secondary particles in cosmic ray collisions with ambient matter. As possible production regions they consider matter-accreting condensed objects. Excessive gamma ray production from π^0 decay is avoided in their scenario by having the surrounding gamma ray density large so that gamma-gamma collisions make it optically thick for gamma rays. Antiprotons might be trapped by magnetic fields, but antineutrons could escape from the region and then decay into antiprotons thus providing for their injection into the interstellar medium.

B. More Exotic Explanations

1. Primordial Black Holes and Their Evaporation

Kiraly et al. [1981] and Turner [1983] consider a model involving evaporating primeval black holes (PBH) in the galaxy, first suggested by Hawking [1974]. PBH's with original masses $\sim 5 \times 10^{14}$ g, if created in the early universe, would have evaporated already. Higher mass black holes evaporating and losing mass could contribute to a quasi-equilibrium density of black holes of mass 5×10^{14} g which might contribute to antiprotons observed. Following Carr [1975], Kiraly et al. show that the solar demodulated antiproton spectrum could be consistent in slope and intensity with current ideas regarding

black holes. The low energy flux of antiprotons of Buffington, Schindler, and Pennypacker [1981b] has been demodulated by Kiraly et al. assuming adiabatic energy losses of 400, 600, and 900 MeV, respectively, as shown in Figure 2. This $\overline{P}$ spectrum from primordial black holes would be a power law ($E^{-3.0}$). This type of a power law spectrum for antiprotons could arise either from acceleration of nearly thermal antiprotons or emission from primordial black holes. The acceleration of antiprotons from nearly thermal antiprotons from the galaxy is ruled out because the annihilation gamma ray background would be much larger than observed. The black hole model according to the authors is consistent with the gamma ray background, the electron and positron fluxes, and the low abundance of antihelium. Kiraly et al. also pointed out that if black holes are confined to galaxies, the antiproton data reduce their upper limit by a factor of 30 compared to the present limits set by gamma ray background.

2. Galactic Nuclei, SS433 Type Objects and Their Environment

Eichler [1982] points out that while solar modulation may enhance $\overline{P}/P$ at low energies, the modulation effect alone is not strong enough to account for the value claimed by Buffington, Schindler, and Pennypacker [1981b]. In solar modulation, while there is energy loss, the intensity of antiprotons would be even higher outside the heliosphere than the observed values. According to Eichler, antiprotons could be produced in dense compact regions where the radiation density is sufficiently high to block gamma ray escape through photon-photon collisions and to degrade electron energies. Adiabatic deceleration of the produced antiprotons is postulated. Following Ramaty and Lingenfelter [1981], Eichler suggests that the environment around objects such as active galactic nuclei or those like SS433 may be suitable candidate sources for injecting antiprotons into the interstellar medium.

3. Baryon Symmetric Cosmologies

The observed particle/antiparticle symmetry in accelerator experiments and the conservation of baryon and lepton numbers in particle interactions leads one to question why this symmetry is not observed in the universe. Several

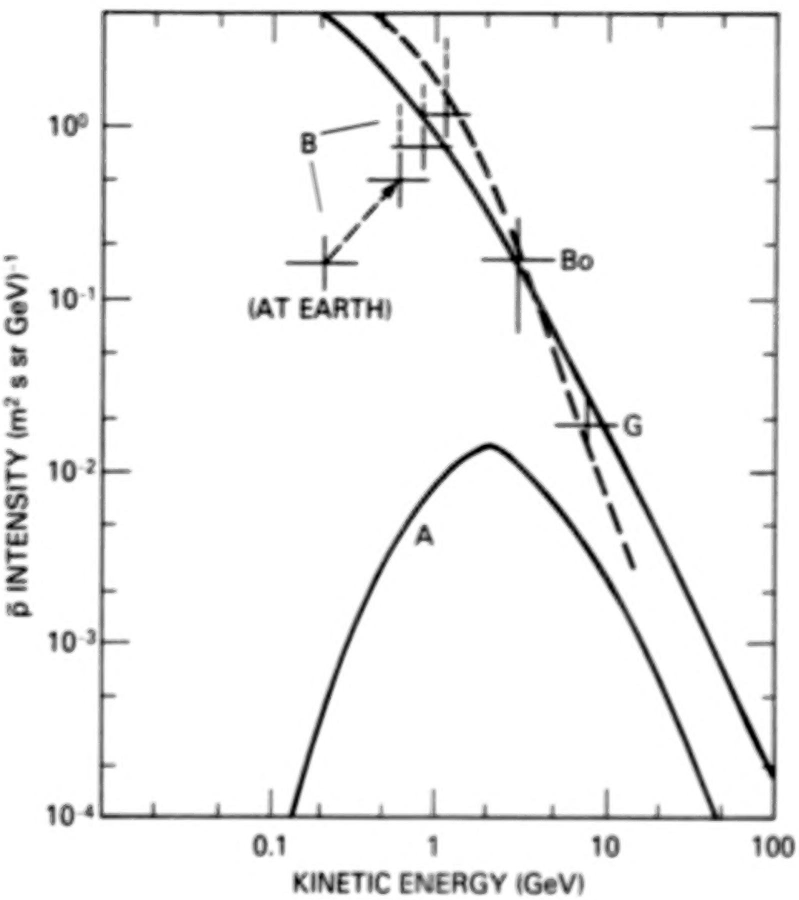


Figure 2. Following Kiraly et al. [1981] and Stecker and Wolfendale [1984] the data are shown with antiproton spectra expected from primordial black holes (dashed curve) and extragalactic antiprotons (solid curve). The differential flux represented by the [Buffington, Schindler, and Pennypacker, 1981b] point has been demodulated assuming three different mean energy losses (see text), adiabatic deceleration, and the applicability of Liouville's theorem. The vertical dashes on the three demodulated points represent plausible uncertainties in this procedure. Both spectra have been normalized at 9 GeV. For comparison, curve A shows the spectrum expected from a leaky box model with 5 g/cm².

cosmologists have taken the view that the symmetry is a fundamental one; but the separation of matter and antimatter into different regimes prevents the total annihilation of matter. Among the early models, we refer to those developed by Omnes [1970] and studied by Stecker and Puget [1972] and Combes, Fassi-Fahri, and Leroy [1975].

The development of the Grand Unified theories has resulted in the development of models [Stecker, 1982] where the exact manner in which charge parity (CP) violation is incorporated in gauge theories determines the nature of the resulting cosmology. If CP violation is spontaneous, random sign changes in casually independent regions will split the universe into domains of baryon or antibaryon excesses. Stecker has used this scenario to postulate an explanation of the cosmic gamma ray background spectrum. If these regions are separated on a galactic scale, a small flux of extragalactic $\overline{P}$ and $\overline{He}$ could be expected.

Stecker, Protheroe, and Kazanas [1983] examined the possibility that the observed antiprotons are primary particles from active galaxies and found this hypothesis consistent with the antiproton observations. Baryon symmetry would naturally seem to result in $\overline{P}/P = \overline{He}/He$; whereas, the results of Buffington, Schindler, and Pennypacker [1981b] imply $\overline{He}/He < \overline{P}/P$ by at least a factor of 10. This result requires destruction of antihelium relative to antiprotons. This is done by fragmentation loss of antihelium by interactions with matter or radiation. Stecker et al. presume that the $\overline{He}$ produced in active galaxies is destroyed, so that any $\overline{He}$ surviving would come from normal (anti) galaxies. Estimating leakage from such galaxies, they predict $\overline{He}/He \sim 5 \times 10^{-6}$ to 5×10^{-5}, close to the Buffington, Schindler, and Pennypacker limit of 2×10^{-5}.

More stringent experimental limits on $\overline{He}/He$ would be of great value in the context of baryon symmetric models.

Stephens [1983] observed that the high energy $\overline{P}/P$ ratio in our galaxy should vary as E^{δ}, when the leakage of cosmic rays from galaxies is assumed to vary as a power law $E^{-\delta}$. This assumes: (1) that the source spectra in our galaxy and in an extragalactic source are the same and (2) that the source spectral index of protons in our galaxy is harder than that observed by an amount

δ. As a consequence, extragalactic antiprotons arrive at Earth with a spectrum $E^{-2.7+\delta}$ and the $\overline{P}/P$ ratio observed rises as E^{δ}. (See Figure 3.) Stecker and Wolfendale [1984] then suggest that the increasing flux of extragalactic antiprotons and protons might account for the "bump" in the cosmic ray spectrum observed around 10^{15} eV.

Stephens [1983, 1985] has examined the question of constraints on the high energy antiproton spectrum derived from the observed sea level muon charge ratio. He derives upper limits on the $\overline{P}/P$ ratio such that the extragalactic hypothesis would conflict with values of δ greater than about 0.6 for energies above 10^4 GeV.

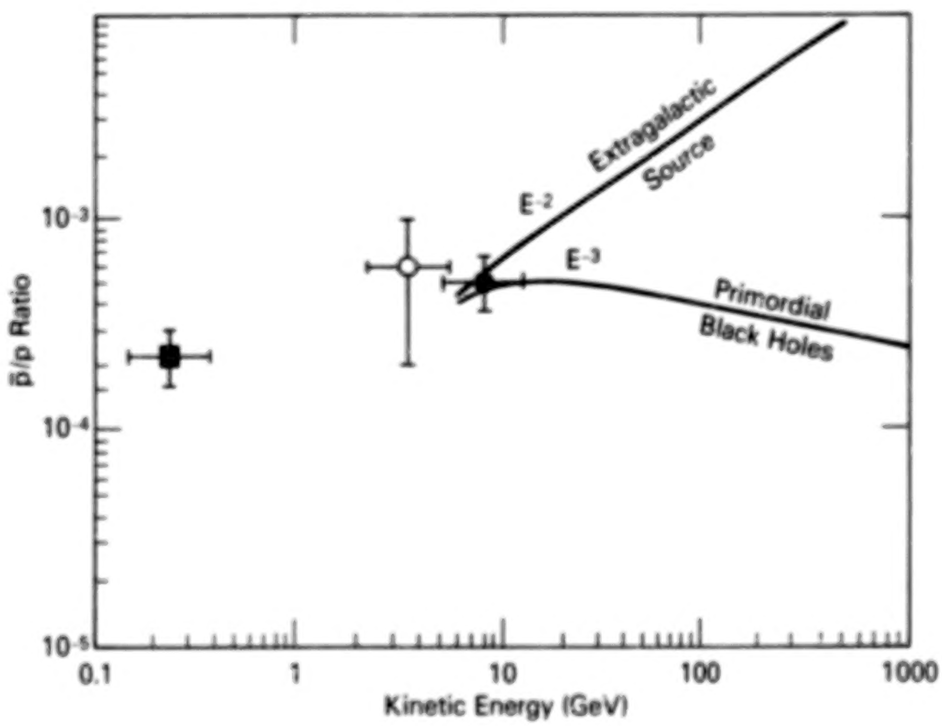


Figure 3. The expected high energy behavior of antiproton spectra from extragalactic sources and primordial black holes are shown.

Measurement of the $\overline{P}$ spectrum above 10 GeV would either decisively rule out the extragalactic origin of antiprotons or provide evidence for its validity.

4. Supersymmetric Theories and Photinos

In supersymmetric theories, for every boson there is a corresponding fermion partner. Silk and Srednicki [1984] suggested that the photino, the supersymmetric partner of the photon, could be a candidate for the invisible mass in the universe and could help in a viable scenario for galaxy formation and clustering of galaxies. The signature of the presence of vast amounts of photinos in the early universe could be looked for in the larger flux of antiprotons produced in the annihilation of photinos and antiphotinos. Stecker, Rudaz, and Walsh [1985] calculate that the available data on antiprotons are consistent with photinos of several GeV mass, Figure 4. Their calculations are normalized to fit the existing data, but the shape of the $\overline{P}/P$ ratio is calculated for the photino masses indicated. Assuming all the observed antiprotons are due to this process, and ignoring for the moment the large uncertainties in the existing experimental data, Stecker et al.'s calculation seems to suggest a photino mass of 15 GeV or higher. More precise $\overline{P}$ data which showed a sharp cutoff in the antiproton spectrum would be strong evidence for photinos or other Majorana fermions in the galaxy. Some recent calculations indicate that gamma ray lines may also be a signature of these photinos [Srednicki, Theisen, and Silk, 1986; Rudaz, 1986; and Eichler and Adams, 1987].

4. CONCLUSION

From the summary presented, we can see that the approximately 50 antiprotons collected in balloon experiments to date have generated considerable theoretical interest. Clearly, confirmatory experiments and measurements over an extended energy range are required before definite conclusions are drawn. We can see that antiproton measurements have a bearing on astrophysical problems ranging from cosmic ray propagation to issues of cosmological import. The next generation of balloon experiments and the Particle Astrophysics Magnet Facility being discussed for operation on NASA's Space Station should provide data and new insights of the highest interest.

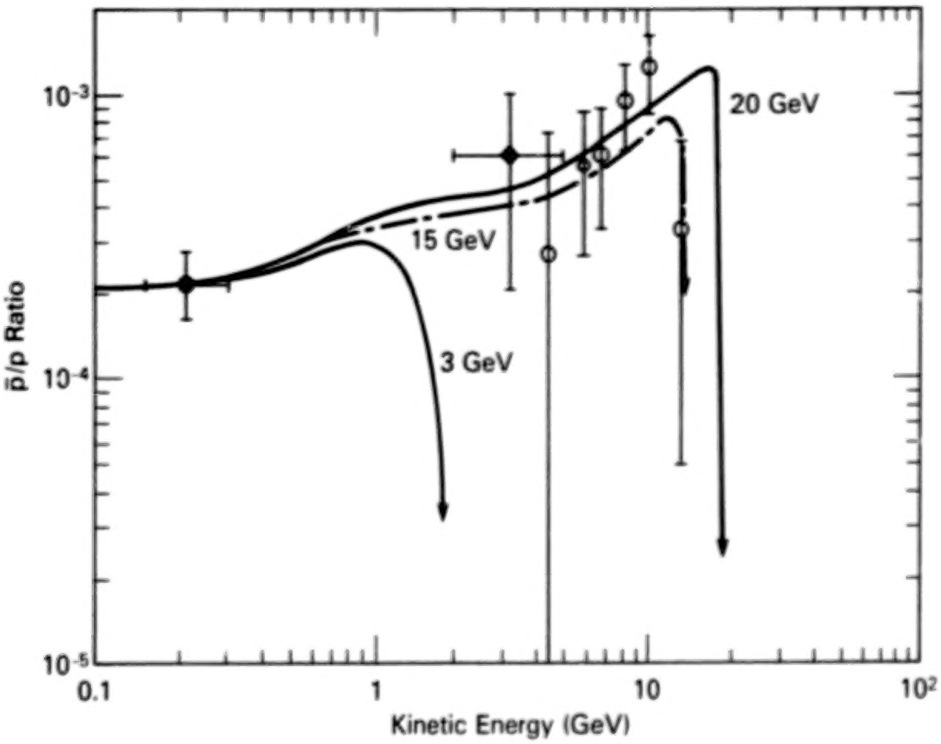


Figure 4. Stecker, Rudaz, and Walsh [1985] compare $\overline{P}/P$ ratios predicted from photino annihilation in the galactic halo with the data, including Golden et al.'s 1984 presentation of his data (open circles). Curves are shown for assumed photino masses of 3, 15, and 20 GeV.

REFERENCES

Anderson, C. D., 1933, *Phys. Rev.*, **43**, 491.

Aizu, H. et al., 1961, *Phys. Rev.*, **121**, 1206.

Apparao, M. V. K., 1967, *Nature*, **215**, 727.

Bogomolov, E. A., Lubyanaya, N. D., Romanov, V. A., 1971, *12th Internat. Cosmic Ray Conference Papers (Tasmania)*, **5**, 1730.

Bogomolov, E. A., Lubyanaya, N. D., Romanov, V. A., Stepanov, S. V., and Shulakova, M. S., 1979, *16th Internat. Cosmic Ray Conference Papers (Kyoto)*, **1**, 330.

Buffington, A., and Schindler, S. M., 1981a, *Astrophys. J.*, **247**, L105.

Buffington, A., Schindler, S. M., and Pennypacker, C. R., 1981b, *Astrophys. J.*, **248**, 1179.

Carr, B. J., 1975, *Astrophys. J.*, **201**, 1.

Cesarsky, C. J., 1980, *Ann Rev. Astron. Astrophys.*, **18**, 289.

Cesarsky, C. J., and Ormes, J. F., 1986, this volume.

Chamberlain, O. et al., 1955, *Phys. Rev.*, **100**, 947.

Combes, F., Fassi-Fahri, O., and Leroy, B., 1975, *Nature*, **253**, 25.

Cowsik, R., and Gaisser, T. K., 1981, *17th Internat. Cosmic Ray Conference Papers (Paris)*, **2**, 218.

Damle, S., Yashdal, V., Rengarajan, T. N., Tandon, S. N., and Verma, R. P., 1973, *13th Internat. Cosmic Ray Conference Papers* **1**, 231.

Dermer, C. D., and Ramaty, R., 1985, *Nature*, **319**, 205.

Dirac, P. A., 1928, *Proc. Roy. Soc., A.*, **117**, 610.

Durgaprasad, N., and Kunte, P. K., 1971, *Nature*, **234**, 74.

Eichler, D., 1982, *Nature*, **295**, 391.

Eichler, D., and Adams, J., 1987, *Astrophys. J.*, in press.

Evenson, P., 1972, *Astrophys. J.*, **176**, 797.

Gaisser, T. K., and Maurer, R. H., 1973, *Phys. Rev. Letters*, **30**, 1264.

Ginzburg, V. L., and Ptuskin, V. S., 1981, *Sov. Astr. Letters*, **7**, 325.

Golden, R. L. et al., 1979, *Phys. Rev. Letters*, **43**, 1196.

Golden, R. L., Mauger, B. G., Nunn, S., and Horan, S., 1984, *Astrophys. Letters*, **24**, 75.

Hawking, S. W., 1974, *Nature*, **293**, 120.

Kiraly, P., Szabelski, J., Wdowczyk, J., and Wolfendale, A. W., 1981, *Nature*, **293**, 120.

Lagage, P. O., and Cesarsky, C. J., 1985, *Astron. Astrophys.*, **147**, 127.

Lebrun, F. et al., 1983, *Astrophys. J.*, **274**, 231.

Mauger, B. G., and Stephens, S. A., 1983, *18th Internat. Cosmic Ray Conference Papers (Bangalore)*, **2**, 95.

Moraal, H., and Axford, W. I., 1983, *Astron Astrophys.*, **125**, 204.

Morfill, G. E., Meyer, P., and Lust, R., 1985, *Astrophys. J.*, **296**, 670.

Omnes, R., 1970, *Phys. Reports C*, **3**, 1.

Peters, B., and Westergaard, N. J., 1977, *Ap. Sp. Sci.*, **48**, 21.

Protheroe, R. J., 1981, *Astrophys. J.*, **251**, 287.

Ramaty, R., and Lingenfelter, R. E., 1981, *Phil. Trans. Roy. Soc., A*, **301**, 671.

Rudaz, S., 1986, *Phys. Rev. Letters*, **56**, 2128.

Silk, J., and Srednicki, M., 1984, *Phys. Rev. Letters*, **53**, 624.

Smoot, G. F., Buffington, A., and Orth, C. D., 1975, *Phys. Rev. Letters*, **35**, 285.

Srednicki, M., Theisen, S., and Silk, J., 1986, *Phys. Rev. Letters*, **56**, 263.

Stecker, F. W., and Wolfendale, A. W., 1984, *Nature*, **309**, 37.

Stecker, F. W., and Puget, L., 1972, *Astrophys. J.*, **178**, 57.

Stecker, F. W., 1982, in *Progress in Cosmology*, ed. A. W. Wolfendale (Dordrecht: D. Reidel Publishing Co.), p. 1.

Stecker, F. W., 1983, *IAU Symp.*, **104**, 437.

Stecker, F. W., Protheroe, R. J., and Kazanas, D., 1983, *Astrophys. and Space Sci.*, **96**, 171.

Stecker, F. W., Rudaz, S., and Walsh, T. F., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **2**, 358.

Steigman, G., 1976, *Ann. Rev. Astr. and Astrophys.*, **14**, 379.

Stephens, S. A., 1981a, *17th Internat. Cosmic Ray Conference Papers (Paris)*, **13**, 89.

Stephens, S. A., 1983, *18th Internat. Cosmic Ray Conference Papers (Bangalore)*, **9**, 167.

Stephens, S. A., 1985, *Astron. Astrophys.*, **149**, 1.

Stephens, S. A., 1981, *Nature*, **289**, 267.

Tan, L. C., and Ng, L. K., 1983, *Astrophys. J.*, **269**, 751.

Turner, M. S., 1983, *Nature*, **297**, 379.

Zeldovich, Y. B., 1965, *Adv. Astron. Astrophys.*, **3**, 241

N87
24256
UNCLAS

N87-24256

9

MEASUREMENTS OF ULTRAHEAVY COSMIC RAYS WITH HEAO-3

W. R. Binns, M. H. Israel, J. Klarmann
Department of Physics and the McDonnell Center for the Space Sciences
Washington University
St. Louis, Missouri 63130

T. L. Garrard, E. C. Stone
George W. Downs Laboratory
California Institute of Technology
Pasadena, California 91125

C. J. Waddington
School of Physics and Astronomy
University of Minnesota
Minneapolis, Minnesota 55455

ABSTRACT

The HEAO-3 Heavy Nuclei Experiment has measured abundances of elements from $_{18}$Ar to $_{92}$U in the cosmic rays. The results on the ultraheavy elements, those with atomic number greater than 30, indicate that the sources of cosmic rays contain a mixture of r-process and s-process material similar to that found in the solar system. This result is at variance with previous indications that the sources are greatly enhanced with freshly synthesized r-process material. Apparent discrepancies between our results and the accepted solar-system abundances have led to a reexamination of data on photospheric abundances of Ge and Pb, resulting in suggested reductions in their values.

1. INTRODUCTION

It is appropriate to discuss the ultraheavy (UH) cosmic rays (nuclei with atomic number, Z, greater than 30) in a volume which honors the career of Frank McDonald. While Frank has not participated directly in experiments to measure these very rare nuclei, he has had major influence on work in this field, both by his pioneering research in the mid-1950's with multiparameter counter techniques for identifying cosmic ray elements, and by his leadership in the late 1960's and the 1970's in establishing and guiding the High Energy Astronomy Observatory (HEAO) program. The best data to date on the elemental composition of UH cosmic rays has come from the Heavy Nuclei Experiment which flew on HEAO-3, and this instrument used a multiparameter counter technique which is a direct descendent of that used in Frank's early work.

The first counter telescope which used the dE/dx-Cherenkov technique was flown by Frank on balloons in 1955 [McDonald, 1956]. In those flights he used an NaI scintillation counter and a Lucite Cherenkov counter to measure the cosmic ray alpha-particle energy spectrum. Subsequently Frank and Bill Webber extended the use of this technique in an important series of balloon flights which measured the proton and alpha-particle energy spectra and their variation over the course of several years [McDonald and Webber, 1959, 1960].

In the late 1960's Frank was the principal motivator behind plans for a "Super Explorer" program in which a new class of very large instruments for high energy astrophysics could be placed in orbit. His efforts led in 1970 to a solicitation for proposals for experiments to be flown on a series of High Energy Astronomy Observations. In 1971, a number of X-ray, gamma ray, and cosmic ray experiments were selected for two large HEAO spacecraft. Our Heavy Nuclei Experiment was among those selected for the first HEAO, which at that time was scheduled for launch in 1975. In early 1973, impelled by budget problems in NASA, the HEAO program was reconfigured to three smaller spacecraft, and our experiment was moved to the third of these, scheduled for launch in 1979. As HEAO project scientist, Frank played a major role in maintaining the scientific viability of the HEAO program in the face of these difficult redesigns.

HEAO-3 was launched on September 20, 1979, into a circular orbit with initial altitude 496 km and inclination 43.6°. The spacecraft returned data until the end of May 1981. The Heavy Nuclei Experiment [Binns et al., 1981] was composed of six dual-gap, parallel-plate pulse ionization chambers, a Cherenkov counter with Pilot-425 (plexiglass doped with wavelength shifter) radiators, and four layers of dual-coordinate multiwire hodoscopes. The instrument used the dE/dx-Cherenkov technique for measuring the nuclear charge of individual elements. The total geometry factor of the HEAO Heavy Nuclei Experiment was approximately 5 m^2sr, although the best charge resolution was achieved by limiting analysis to particles that penetrated all the counters—a geometry factor of approximately 1 m^2sr. This instrument achieved individual-element resolution for even-Z elements from $_{26}$Fe through $_{58}$Ce, and achieved adequate resolution at higher atomic numbers to determine the ratio of the Pb-group to the Pt-group of elements and the relative abundance of actinide elements.

Preliminary results from this experiment were reviewed at the International Cosmic Ray Conference in Paris [Israel, 1981] and in the proceedings of the 1982 summer course in Erice [Israel, 1983a]. In Erice we also presented a summary of UH detector techniques [Israel, 1983b]. A later review appeared in a 1984 Committee on Space Research (COSPAR) symposium [Binns et al., 1984], and a more exhaustive review of UH cosmic rays is in preparation.

In this paper we summarize our results thus far which have a bearing on the elemental composition at the cosmic ray source. The objective is to compare the observed cosmic ray abundances with those expected from various plausible compositions at the cosmic ray source, and from the comparison to study the nucleosynthesis history of these cosmic rays as well as elemental fractionation effects which may occur in the acceleration process. In so doing we concentrate on those elements whose observed abundances at Earth are unlikely to include a very large component of fragments from the collisions of heavier cosmic rays with nuclei of the interstellar gas. For these elements, calculations which account for the interstellar fragmentation do not depend very sensitively upon the details of the model of galactic confinements of these nuclei, although some propagation calculations are essential to this analysis [Brewster, Freier, and Waddington, 1983, 1985; Margolis and Blake, 1983, 1985].

Work on other aspects of this Heavy Nuclei Experiment is in progress, and preliminary results have been reported elsewhere. The abundances of secondary UH elements were discussed by Klarmann et al. [1985]. We have reported results from calibrations of the instrument with relativistic heavy ions from the Bevalac at the Lawrence Berkeley Laboratory [Newport et al., 1985] and preliminary results from studies of heavy ion fragmentation carried out in connection with this Bevalac calibration [Kertzman et al., 1985]. This experiment has also allowed us to measure the relative abundance of Fe and Fe-secondary elements at energies up to several hundred GeV/amu [Jones et al., 1985].

There are four groups of mainly-primary UH elements: $32 \leq Z \leq 42$, $50 \leq Z \leq 58$, $76 \leq Z \leq 82$, and $Z \geq 90$. In the following section of this paper we summarize our results on each of these groups in turn.

2. RESULTS

Prior to the results of this experiment, observations using passive detectors (nuclear emulsions and plastic track detectors) had indicated that the cosmic rays were greatly enriched in elements produced by r-process nucleosynthesis. The very large values found for the ratio of actinide elements ($Z>89$) to elements of the platinum-lead group ($74 \leq Z<84$), 10% or more, compared with the 1% expected from a cosmic ray source with solar system abundances, implied that the cosmic rays were significantly enriched in freshly synthesized r-process elements [Fowler et al., 1977; Shirk and Price, 1978]. Such enrichment would be expected if supernovae supply the energy for the cosmic rays and accelerate material from regions where r-process nucleosynthesis is taking place.

One of the principal achievements of this experiment has been the demonstration that this view is incorrect. We have found element abundances which are remarkably similar to those expected from a source with composition very similar to that of the solar system, when effects of fractionation dependent upon first ionization potential are taken into account. Indeed the similarity is so striking that when our results for two elements (Ge and Pb) failed to fit with accepted solar system abundances, investigators were stimulated to

reexamine those abundances, concluding that photospheric abundances of those two elements are actually about a factor or two lower than previously accepted values [Grevesse and Meyer, 1985].

Our observed abundances of elements from $_{32}$Ge through $_{42}$Mo relative to Fe are displayed as data points in Figure 1 [Binns et al., 1984]. The histograms in this figure compare the data with that expected from various source abundances. In each panel the solid line histogram assumes no elemental fractionation at the source, while the dashed line histogram assumes fractionation which depends exponentially on the element's first ionization potential (FIP). There is reasonably good agreement between the data and the prediction from solar system abundances [Cameron, 1982a] with FIP fractionation. Since the solar system abundances for these elements are dominated by s-process nucleosynthesis [Cameron, 1982b], there is also reasonable agreement between the data and the s-process prediction. The abundances of these elements relative to Fe are much greater than those of the r-process component of the solar system, and even if one renormalizes the r-process abundances, the pattern of element-to-element abundances variations does not match the data as well as a simple solar system source.

One notable exception to the agreement between our data and the solar system abundances is the element Ge. This element is well resolved in our data, and its abundance relative to Fe is about half that which would be expected from the solar system source. A similar conclusion is reached by examining the data from the other HEAO-3 cosmic ray experiment, although that experiment has a smaller geometry factor and thus its conclusion about Ge abundance has lower statistical significance [Byrnak et al., 1983]. This conclusion is the same whether one compares the cosmic ray data with the Cameron [1982a] solar system abundances or those of Anders and Ebihara [1982]; and it is unaffected by any model of source fractionation in which first ionization potential is the organizing parameter, because Fe and Ge have almost exactly the same value of FIP. We have previously noted [Israel et al., 1983] that the low observed Ge abundance could be explained if volatility were a significant factor in source fractionation, as had been suggested by several authors [Cesarsky and Bibring, 1980; Epstein, 1980; Bibring and Cesarsky, 1981, Tarafdar and Apparao, 1981; and Meyer, 1981]. We also noted [Binns et al., 1984]

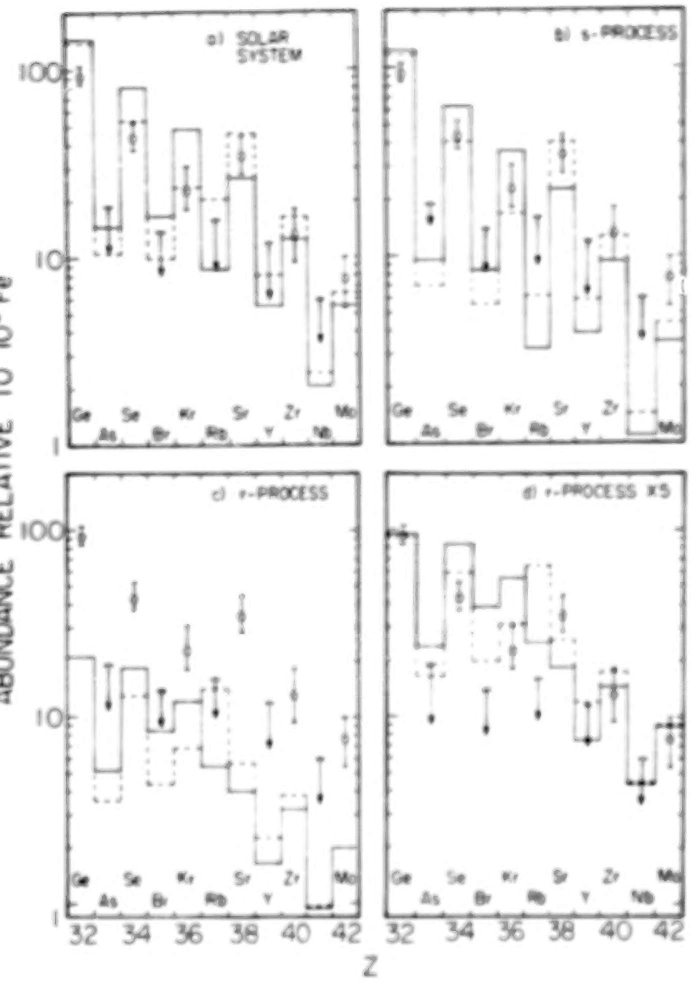


Figure 1. Comparison of our measured abundances (data points) [Binns et al., 1983] with those expected at Earth after galactic propagation, through an exponential path length distribution with mean 5.5 g/cm² of hydrogen, from a source with (a) solar system abundances [Cameron, 1982a], (b) solar system s-process abundances [Cameron, 1982b], (c) solar system r-process abundances [Cameron, 1982b], and (d) solar system r-process abundances enhanced by a factor of 5. In each panel the solid line assumes no FIP fractionation and the dashed line assumes FIP fractionation of the form 9.31 exp (−0.288FIP).

that the low Ge abundance could indicate that there may be a difference between the Cl-meteorite abundance of this element and the correct solar system abundance.

In the face of this discrepancy between our Ge data and the standard solar system abundances, Grevesse and Meyer [1985] reexamined the spectroscopic data on the photospheric abundance of Ge. They concluded that the best estimate of the photospheric Ge abundance is lower than the Cl meteorite value (and the previously accepted photospheric value) by nearly a factor of 2. When they use this new estimate of the photospheric abundance of Ge as the solar system value, the discrepancy between cosmic ray and solar system Ge abundances disappears.

For the elements $_{50}$Sn through $_{58}$Ce, Figure 2 [Stone et al., 1983] compares measured abundances (data points) with abundances expected (histograms) from various sources. In this charge interval the solar system abundances have about equal overall contributions from r-process and s-process nucleosynthesis, with the r-process dominating the production of $_{52}$Te and $_{54}$Xe and the s-process dominating $_{50}$Sn, $_{56}$Ba, and $_{58}$Ce. If one ignores the possibility of source fractionation dependent on the first ionization potential, then the observed peaks of Sn and Ba suggest a distinct enhancement of s-process material. But these two s-process elements also have lower FIP than the Te and Xe, and thus FIP fractionation would be expected to increase the abundance of Sn and Ba relative to Te and Xe. When FIP fractionation similar to that found for lower-Z elements is applied to the possible sources, the data are found to be in reasonable agreement with a source abundance containing a mixture of r-process and s-process contributions in about the same proportions as is found in the solar system.

In the "platinum-lead" region the solar system abundances are dominated by an r-process peak of the elements $_{76}$Os, $_{77}$Ir, and $_{78}$Pt, and an s-process peak at the element $_{82}$Pb. In our experiment we were unable to resolve individual element peaks at these high charges, but we did form a "Pb/Pt" ratio of charge groups with the "Pb" group including events with charge $81 \leq Z \leq 86$, and the "Pt" group including events with charge $74 \leq Z \leq 80$. We find a value of 0.25 ± 0.09 for this ratio [Binns et al., 1985]. Figure 3 compares this result with that expected from a source with standard solar system

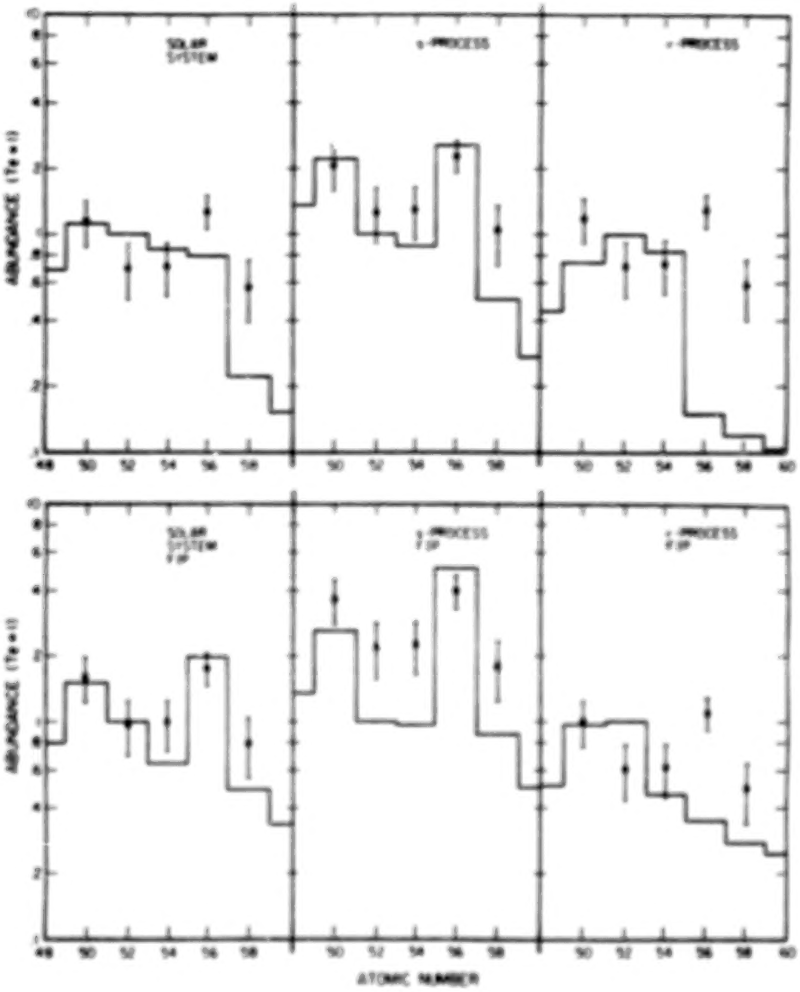


Figure 2. Comparison of our measured abundances (data points) with those expected at Earth after galactic propagation, through an exponential path length distribution with mean 5.5 g/cm² from sources with solar system abundances [Anders and Ebihara, 1982], s-process abundances [Kaeppeler et al., 1982], or r-process abundances [Krombel, 1983]. The upper panels assume no FIP fractionation and the lower panels assume exponential FIP fractionation of the form 9.31 exp (−0.288FIP).

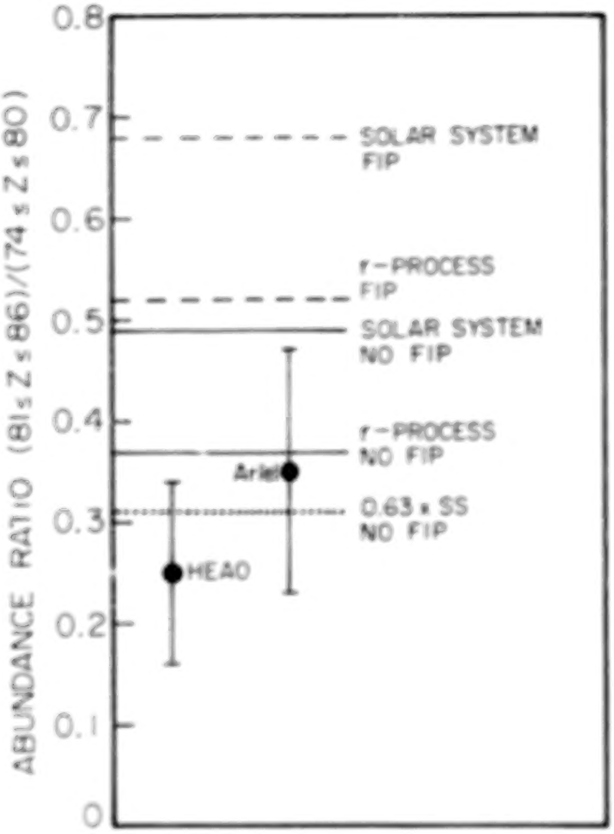


Figure 3. The ratio of "Pb" (81≤Z≤86) to "Pt" (74≤Z≤80). The data points are observations from HEAO-3 [Binns et al., 1985] and Ariel-6 [Fowler et al., 1985]. The lines are expected values after propagation through an exponential path length distribution with mean 5.5 g/cm² from sources with standard solar system abundances [Anders and Ebihara, 1982] or r-process abundances derived from those solar system abundances [Fixsen, 1985; Binns et al., 1985]. Solid lines assume no FIP fractionation, which for these elements is equivalent to a step function FIP fractionation with step above 9 eV; dashed lines assume exponential FIP fractionation. Reduction of the solar system Pb abundances as suggested by Grevesse and Meyer [1985] would lower the expected solar system values by a factor 0.63, as indicated by the dotted line, and would cause an even greater reduction in the r-process expectations.

abundances [Anders and Ebihara, 1982] and with an r-process source derived from those solar system abundances [Fixsen, 1985; Binns et al., 1985]. Our observed "Pb/Pt" ratio is distinctly lower than that expected from this assumed solar system source, suggesting an enhancement in r-process contribution to the cosmic ray source. The UH experiment on the Ariel-6 spacecraft [Fowler et al., 1985] found a "Pb/Pt" ratio, 0.35 ± 0.12, which is consistent with our result.

As with the low observed Ge abundance, the low Pb abundance would be consistent with a volatility dependence of the source fractionation, without invoking a nonsolar system source abundance. However, here too Grevesse and Meyer [1985] were stimulated by our cosmic ray measurement to reexamine the spectroscopic data on the photospheric abundance of Pb. They conclude that the best photospheric abundance for Pb is about 0.63 of the standard (Cl meteorite) abundance. If the "expected" solar system values in Figure 3 are multiplied by 0.63, then our observed value of this ratio is no longer significantly lower than that expected from a solar system source. Thus the observed Pb/Pt ratio would not require a significant enhancement of r-process material.

Finally, we turn to the result on the heaviest elements, the actinides, $Z \geq 90$, summarized in Figure 4. Prior to our HEAO-3 and the Ariel-6 experiments, measurements with nuclear emulsions and plastic track detectors had indicated that in the cosmic rays the ratio of actinides to elements in the "Pb/Pt" region was at least an order of magnitude higher than in the solar system [Fowler et al. 1977; Shirk and Price, 1978]. Since the actinides are produced only by r-process nucleosynthesis, this actinide enrichment would have implied a very significant enrichment of freshly synthesized r-process material in the cosmic ray source.

The actinide abundances reported by Fowler et al. and by Shirk and Price were questioned by Meyer [1979] who concluded from an examination of their data that the evidence for such high abundances was not convincing. Resolution of these conflicting interpretations of the data did not come until the HEAO-3 and Ariel-6 data had been analyzed and O'Sullivan [1985] had reevaluated the earlier balloon data. He concluded, in the light of new understanding of the temperature dependence of track registration in plastics, that

C-3

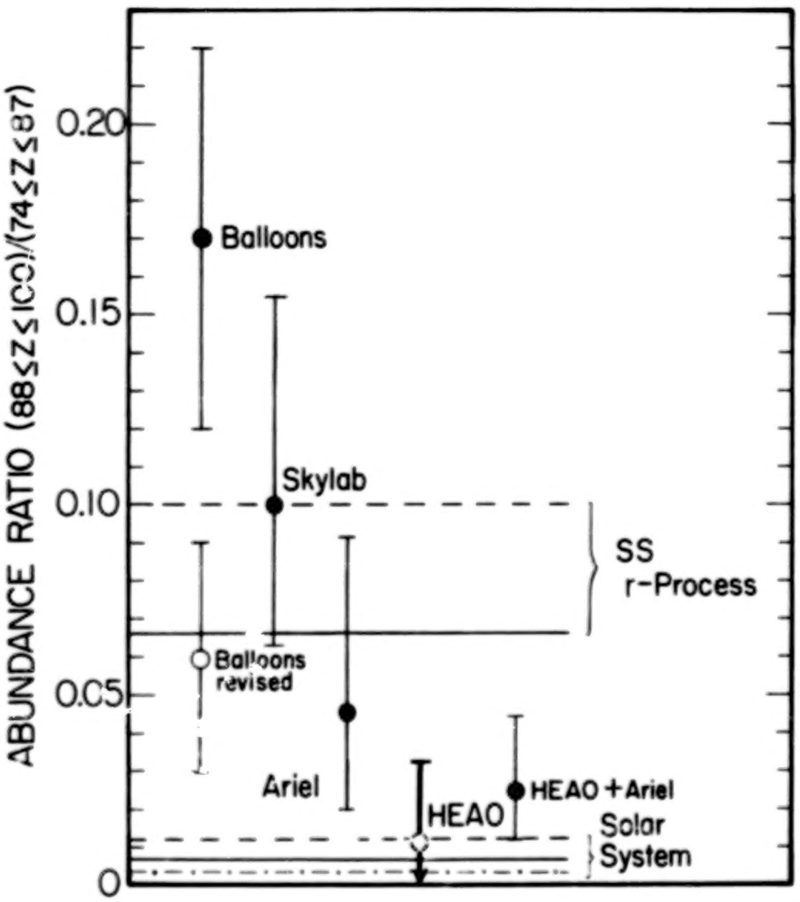


Figure 4. The abundances of "actinides": (88≤Z≤100) relative to "Pt-Pb" (74≤Z≤87). Data are from balloons [Fowler et al., 1977], Skylab [Shirk and Price, 1978], Ariel-6 [Fowler et al., 1985], and HEAO-3 [Binns et al., 1985]. The revised balloon point is from O'Sullivan [1985]. Expected values [Blake et al., 1978] are solid lines without FIP fractionation and dashed lines with FIP fractionation, for sources with composition of the solar system at the time of its formation or for a source with the composition of freshly synthesized r-process material. The dot-dash solar system line substitutes present-day abundances for the (solid line) abundances at the formation of the solar system.

the plastic track balloon data were consistent with an actinide abundance substantially lower than that previously derived from those data.

By carefully examining all the high-charge events in our HEAO-3 data for which a reasonably accurate charge could be assigned, we found just one event which might be an actinide [Binns et al., 1982]. The assigned charge for this event was 89, but the very short half-lives of all the elements in the interval $84 \leqslant Z \leqslant 89$, combined with the poor resolution of this data set, makes it more probable that this was a nucleus of $_{90}$Th or $_{92}$U. At the same time, we cannot be certain that this single event was not in fact a $_{82}$Pb nucleus whose charge was significantly overestimated. Together with this event we found 101 events with $74 \leqslant Z \leqslant 87$, giving us an actinide to "Pt-Pb" ratio of about 1%, with an 84% confidence upper limit of 3%. The Ariel-6 result for the same ratio, based on three "actinide candidates" and 65 in the "Pt-Pb" group [Fowler et al., 1985] is 4.6% (+4.5%/−2.5%). With the very low actinide statistics in these experiments, both the HEAO-3 and the Ariel-6 results are consistent with a result formed by combining the two, 2.4% (+1.9%/−1.2%).

Since a cosmic ray source with solar system abundances gives about 1% for the expected value of this ratio, the observations are not inconsistent with such a solar system source. Although with the very low observed statistics one cannot rule out a significant enhancement of r-process actinides in the cosmic ray source, we can exclude the possibility that the cosmic rays consist primarily of freshly synthesized r-process material.

3. DISCUSSION

The relative abundances of UH elements at the cosmic ray source appear to be consistent with those expected from source abundances with a mix of r-process and s-process nucleosynthesis similar to the solar system mix, provided one takes into account elemental fractionation dependent upon first ionization potential, similar to the fractionation which has been observed for elements with atomic numbers below 30. However, it is important to recognize that limitations on statistics, particularly at the highest atomic numbers, mean that we cannot exclude the possibility of factor-of-two differences between the r-process/s-process ratio in the solar system and that in the cosmic ray

source. Indeed, when we consider the differences between the isotopic composition of Ne, Mg, and Si at the cosmic ray source and in the solar system [Wiedenbeck, 1984] it would be surprising if the UH cosmic ray source composition did not have some differences of perhaps as much as a factor of two from the solar system.

It had appeared that the cosmic ray abundances of Ge and Pb were low by a factor of about two relative to nearby elements when our measurements were compared with abundances expected from the standard compilations of solar system abundances based on Cl meteorites. But this discrepancy disappeared when data on photospheric abundances were reexamined and the new photospheric values were substituted for the Cl-meteorite values in the compilation of solar system abundances.

The picture that emerges from these new observations of UH cosmic rays is consistent with models of shock acceleration of cosmic rays in the interstellar medium. In these models, the energy of the cosmic rays comes from supernova explosions, but the nuclei themselves come from the interstellar medium. Since the solar system condensed out of interstellar medium, in these models one expects the cosmic ray and solar system abundances to be similar, although one might expect differences in detail owing to the differences in time and place at which these two sets of abundances sample the interstellar medium. Those detailed differences could be of great importance to our understanding of the chemical evolution of the galaxy, but their study awaits future experiments [Drach et al., 1985] with individual element resolution and much larger collecting power.

ACKNOWLEDGMENTS

The success of the HEAO program was the result of cooperation and hard work of many people. In addition to Frank, whose vital role we described in the introduction above, we are pleased to acknowledge the efforts of the spacecraft contractor TRW, the NASA project team at Marshall Space Flight Center, and the NASA program office. We are grateful to the people who were essential to the fabrication of the instrument—the staff at Ball Aerospace, especially Herb Chameroy, Art Gneiser, and Mike Noble; and the people at

our universities, namely, Bill Blodgett, John Epstein, Bill Erickson, Bob Howard, and Bob Scarlett. Several physicists in our laboratories have contributed significantly to the data analysis, and we thank Mary Brewster, Bob Fickle, Dale Fixsen, Mike Jones, Mary Kertzman, Keith Krombel, and Brian Newport. Much valuable computer programming was done by Tom Aufrance, Brownlee Gauld, Dave Grossman, and Marty Olevitch. We are grateful to HEAO Guest Investigators Steve Margolis, whose cosmic ray propagation code and insights into the propagation problem have been very useful in our interpretation of the data, and Peter Eby, whose calculations of expected signals in our detector system were useful in an early stage of our data analysis. We are pleased to acknowledge the excellent cooperation of the staff of the Bevalac at the Lawrence Berkeley Laboratory, and we particularly thank Hank Crawford for his tireless efforts in support of our instrument calibration. During the design, development, and early data analysis of this experiment one of us (WRB) was at the McDonnell Douglas Research Laboratories, and we are pleased to acknowledge the institutional support of MDRL.

This work was supported in part by NASA contracts NAS 8-27976, 77, 78; by NASA grants NAG 8-498, 500, 502; and by NASA grants NGR 05-002-160, 24-005-050, 26-008-001.

REFERENCES

Anders, E., and Ebihara, M., 1982, *Geochim. et Cosmochim.*, **46**, 2363.

Bibring, J. P., and Cesarsky, C. J., 1981, *17th Internat. Cosmic Ray Conference Papers (Paris)*, **2**, 289.

Binns, W. R., Israel, M. H., Klarmann, J., Scarlett, W. R., Stone, E. C., and Waddington, C. J., 1981, *Nucl. Instr. Meth.*, **185**, 415.

Binns, W. R. et al., 1982, *Astrophys. J.*, **261**, L117.

Binns, W. R. et al., 1983, *18th Internat. Cosmic Ray Conference Papers (Bangalore)*, **9**, 106.

Binns, W. R. et al., 1984, *Adv. Space Res.*, **4**, (2-3), 25.

Binns, W. R. et al., 1985, *Astrophys. J.*, **297**, 111.

Blake, J. B., Hainebach, K. L., Schramm, D. N., and Anglin, J. D., 1978, *Astrophys. J.*, **221**, 694.

Brewster, N. R., Freier, P. S., and Waddington, C. J., 1983, *Astrophys. J.*, **264**, 324.

Brewster, N. R., Freier, P. S., and Waddington, C. J., 1985, *Astrophys. J.*, **294**, 419.

Byrnak, B. et al., 1983, *18th Internat. Cosmic Ray Conference Papers (Bangalore)*, **2**, 29.

Cameron, A. G. W., 1982a, in *Essays in Nuclear Astrophysics*, ed. C. A. Barnes, D. D. Clayton, and D. N. Schramm (Cambridge: Cambridge University Press), p. 23.

Cameron, A. G. W., 1982b, *Ap. Space Sci.*, **82**, 123.

Cesarsky, C. J., and Bibring, J. P., 1980, in *IAU Symposium 94, Origin of Cosmic Rays*, ed. G. Setti, G. Spada, and A. W. Wolfendale (Dordrecht: D. Reidel Publishing Co.), p. 361.

Drach, J., Price, P. B., Salamon, M. H., Tarle, G., and Ahlen, S. P., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **2**, 131.

Epstein, R. I., 1980, *M.N.R.A.S.*, **193**, 723.

Fixsen, D. J., 1985, *University of Minnesota Cosmic Ray Report*, CR-195.

Fowler, P. H., Alexandre, C., Clapham, V. M., Henshaw, D. L., O'Sullivan, D., and Thompson, A., 1977, *15th Internat. Cosmic Ray Conference Papers (Plovdiv)*, **11**, 165.

Fowler, P. H., Masheder, M. R. W., Moses, R. T., Walker, R. N. F., Worley, A., and Gay, A. M., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **2**, 119.

Grevesse, N., and Meyer, J. P., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **3**, 5.

Israel, M. H., 1981, *17th Internat. Cosmic Ray Conference Papers (Paris)*, **12**, 53.

Israel, M. H., 1983a,b, in *Composition and Origin of Cosmic Rays*, ed. M. M. Shapiro (Dordrecht: D. Reidel Publishing Co.), pp. 47, 291.

Israel, M. H. et al., 1983, *18th Internat. Cosmic Ray Conference Papers (Bangalore)*, **9**, 305.

Jones, M. D. et al., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **2**, 28.

Kaeppeler, F., Beer, H., Wissak, K., Clayton, D. D., Macklin, R. L., and Ward, R. A., 1982, *Astrophys. J.*, **257**, 821.

Kertzman, M. P. et al., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **3**, 95.

Klarmann, J. et al., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **2**, 127.

Krombel, K. E., "The Relative Abundances of Sn, Te, Xe, Ba, and Ce in the Cosmic Radiation." Ph.D. dissertation, California Institute of Technology, 1983.

Margolis, S. H., and Blake, J. B., 1983, *18th Internat. Cosmic Ray Conference Papers (Bangalore)*, **9**, 283.

Margolis, S. H., and Blake, J. B., 1985, *Astrophys. J.*, **299**, 334.

McDonald, F. B., 1956, *Phys. Rev.*, **104**, 1723.

McDonald, F. B., 1959, *Phys. Rev.*, **116**, 463.

McDonald, F. B., and Webber, W. R., 1959, *Phys. Rev.*, **115**, 194.

McDonald, F. B., and Webber, W. R., 1960, *J. Geophys. Res.*, **65**, 767.

Meyer, J. P., 1979, *16th Internat. Cosmic Ray Conference Papers (Kyoto)*, **1**, 374.

Meyer, J. P., 1981, *17th Internat. Cosmic Ray Conference Papers (Paris)*, **2**, 281.

Newport, B. J. et al., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **3**, 287.

O'Sullivan, D., 1985, *Irish Astron. J.*, **17**, 40.

Shirk, E. K., and Price, P. B., 1978, *Astrophys. J.*, **220**, 719.

Stone, E. C. et al., 1983, *18th Internat. Cosmic Ray Conference Papers (Bangalore)*, **9**, 115.

Tarafdar, S. P., and Apparao, K. M. V., 1981, *17th Internat. Cosmic Ray Conference Papers (Paris)*, **2**, 285.

Wiedenbeck, M. E., 1984, *Adv. Space Res.*, **4**, 15.

N87
24257
UNCLAS

10

ORIGIN AND PROPAGATION OF GALACTIC COSMIC RAYS

Catherine J. Cesarsky
Centre d'Etudes Nuclearires de Saclay
91191 Cedex Gif sur Yvette
France

Jonathan F. Ormes
Laboratory for High Energy Astrophysics
Goddard Space Flight Center
Greenbelt, Maryland 20771

1. INTRODUCTION

In the few years following the launch of HEAO-C with its two large cosmic ray experiments on board, we have seen significant progress made in our understanding of the origin of energetic particles in the galaxy. This progress was made with large, high resolution instruments above the atmosphere for extended periods. It was Frank McDonald's foresight which led to the initiation of the HEAO project and his energy which helped to lead it to a successful conclusion. It is fitting that on the occasion of Frank's sixtieth birthday we should review our understanding of the problems associated with the origin of cosmic rays, problems which have been so central to his scientific interests and to the solution of which he has contributed so much. These contributions have come not only through his own scientific work, but also through his tireless efforts in promoting space flight opportunities and in the development of new scientific talent. This is evidenced by the range of papers in this volume, and by the impact of the HEAO satellites and their experiments on the discipline of high energy astrophysics. In particular, the role played by

HEAO-3 and the Danish-French experiment in furthering our understanding of cosmic rays will be evident in this paper. C. J. Cesarsky introduced the oral presentation of this paper as follows:

> I started working on cosmic rays in 1969. By then, Frank McDonald was already famous, and I was of course very intimidated to meet him at my first colloquium, which was at Goddard in 1971. I was surprised to find that this man was so unassuming and easy to talk to. Over the years, with scientific meetings, and his frequent trips to France where I had located, a friendship developed, based on common interests: cosmic rays, space experiments, good food, and art. So it is a great pleasure to be here on this occasion.

Heavy elements in the galactic cosmic rays were discovered almost forty years ago now [Freier et al., 1948; Bradt and Peters, 1948], and a large number of balloon and satellite observations have been made in the succeeding years. It is rather remarkable that most of these observations can be understood in the framework of a rather simple theory. It is based on the minimum assumption that there is one type of source and one confinement region in which particles are contained by one mechanism. It also assumes that all species, namely electrons, protons, helium, and the heavier elements which we observe are a consequence of the same processes. We will see that recent observations are making this point of view more and more difficult to maintain. This should come as no surprise. For the first time we have highly accurate data—in some cases the principle errors are coming from uncertainties in cross-sections rather than from the cosmic ray data itself. As the level of detail in our observations increases, in effect we are observing the phenomena in "higher and higher resolution". In fact the remarkable thing is the large number of observations which are understood from the perspective of this simple theory.

The Danish-French experiment on HEAO-3 has provided us with our first detailed observations outside the Earth's magnetosphere of particles above 1 GeV/amu. These observations have shaken the simplest interpretations so that we probably cannot even claim to know the spectrum which is produced by the acceleration mechanism(s), much less to understand the mechanism(s).

An accompanying paper [Binns et al., 1987, this volume] discusses the elemental and isotopic abundances and what they can tell us about the mechanisms for nucleogenesis of cosmic rays and the sites in which they reside before acceleration. Much new has been learned here too, but there are many gaps. The mechanism by which the galaxy is able to concentrate so much of its energy resources in so few of its constituents is the problem of the acceleration mechanism. The future will see it approached not only by working our way backward from the observations, but also by working our way forward from what we know about the sites and mechanisms of nucleosynthesis.

This paper will discuss the observations and their interpretation in context of the physical processes involved. Suggestions for future observations which can be used to attempt to resolve the outstanding questions will form the conclusion.

2. GENERAL BACKGROUND

Energetic particles are ubiquitous in astrophysical plasmas. We see them in the solar system as a result of plasma processes wherever there are motions and magnetic fields. They are accelerated in the magnetospheres of the Earth and Jupiter. They are accelerated by the Sun in magnetic fields associated with solar flares. We see synchrotron radiation which tells us that electrons, and by implication nuclei, are being accelerated in supernova remnants, in pulsar magnetospheres, and in quasars. At the same time we observe particles at Earth which are extremely homogeneous in space and time, apparently coming to us from the galaxy at large.

Historically there have been a number of ideas about the site(s) in which the acceleration of cosmic rays takes place: in the galactic magnetic fields, in supernova remnants, in pulsar magnetospheres, etc., but neither the site nor the acceleration mechanism is well understood. Much theoretical work has been done recently on shock acceleration mechanisms, and examples of shock acceleration are known to be at work in the solar system where they can be studied in situ, but whether these mechanisms can operate on a scale sufficient to account for the galactic cosmic rays is still uncertain. More theoretical work

is needed on the transport of particles on a galactic scale. The mechanism must be continuous over at least five or six orders of magnitude, from GeV energies to perhaps 100 or 1000 TeV.

On the other hand objects such as Cygnus X-3 are apparently producing air showers initiated by gamma rays of 1 to 1000 or more TeV energy. There is sufficient power available from this source to fill the galaxy with cosmic rays of 100 TeV or more. Accretion disks and binary stellar systems may be able to accelerate particles too. Any environment involving magnetic fields and motion is a candidate. It may be that a number of different processes accelerate particles which become the cosmic rays observed at Earth.

At energies of 100 TeV and above the cosmic ray air showers are isotropic to a few parts in 10^4 as shown in Figure 1. This implies that the particles are confined in a large column and that particles are not streaming past the solar system at velocities more than a few tens of km/sec. From the radio continuum observations, we know that cosmic ray electrons are present over much of the galaxy and extend beyond the galactic disk into a halo above and below the disk. The radio map of NGC891, an edge on galaxy seen at 21 cm, is shown in Figure 2 superposed on a photograph from the 200-inch telescope from Allen, Baldwin, and Sancisi, 1978. This intensity profile is similar to that which an extragalactic radio astronomer would see if observation were made of our galaxy from a similar perspective. Cosmogenic nucleides in meteorites, nuclei which have been transformed through the bombardment by energetic cosmic ray nuclei during their exposure in space, can be used to estimate the average flux of cosmic rays over their exposure history. This has been done over time scales of 400, 9×10^5, and 10^9 years. These results say that, within a factor of 2, the cosmic ray intensity has been constant over the last billion years. There is some indication that it may have been a factor of two lower on the 10^9 year time scale, and periodic fluctuations of larger amplitude cannot be ruled out. Most of the particles responsible were in the energy range .3 to 3 GeV/amu, so changes in the slope of the well-known observed power law spectrum cannot be ruled out by these observations either.

These considerations led Ginzburg and Syrovatskii [1964] to posit that the galaxy was filled with energetic particles accelerated within the galaxy which

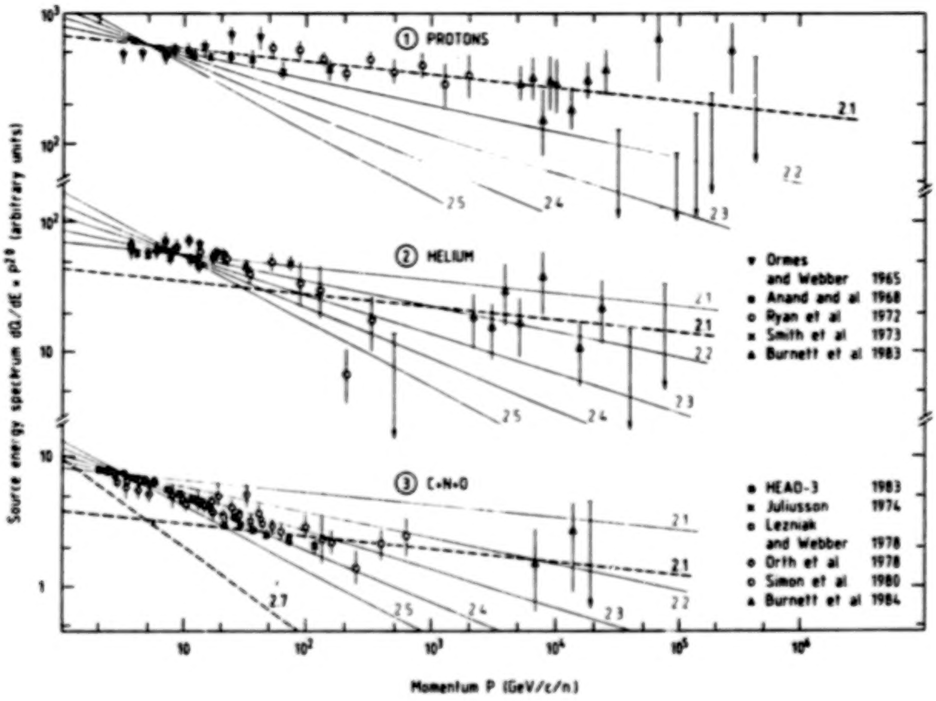


Figure 1. The observed anisotropy is shown as a function of energy. A Compton Getting anisotropy corresponding to a streaming velocity of 20 km/sec is indicated as is the anisotropy which would be expected from a diffusion coefficient varying as the square root of rigidity [from Ormes, 1983, adapted from Hillas, 1984].

diffuse throughout the galactic magnetic fields, thereby remaining trapped for times which are long compared to their straight line travel times across the galaxy. The low anisotropy led them to propose that the galaxy had a halo of turbulent plasma and magnetic fields which acted as the containment volume for cosmic rays. As a result, a steady-state picture arose in which cosmic rays are produced at a given rate and are lost at a given rate, leaving the galaxy

Figure 2. A radio continuum map of the edge on Spiral galaxy NGC891 from the Westerbork Synthesis Radio Telescope at 21.2 cm (1412 MHz) from Allen, Baldwin, and Sancisi, 1978. The contours are shown superposed on a photograph from the 200-inch Palomar telescope courtesy of Hale Observatories.

with a constant or nearly constant density (and intensity) of energetic particles over its lifetime. This led to the phenomenological model we refer to as the "leaky-box" model:

$$Q + \text{spallation} = \text{escape} + \text{interaction} + \text{decay}$$

where the steady injection of particles from sources and the spallation of heavier nuclei to lighter ones is balanced by the loss of particles from the galaxy or their loss due to interaction or decay.

a. The "Leaky-Box" Model

As cosmic ray nuclei spiral through interstellar space, they suffer inelastic collisions with interstellar gas and the "primary" cosmic ray nuclei emitted by sources break up into lighter "secondary" nuclei. The amount of interstellar matter traversed by cosmic rays can be estimated by measuring the abundances of species expected to be rare in the source abundance spectrum. The most prominent of these are lithium, beryllium, and boron, created primarily by the fragmentation of carbon and oxygen nuclei, and the nuclei with atomic numbers 21 to 25, the so-called sub-iron nuclei.

At energies greater than a few GeV/amu, the effects of solar modulation and of Coulomb interactions in the interstellar medium are negligible and the cross-sections of the spallation reactions affecting the cosmic ray composition are nearly energy-independent.

Assuming the interstellar gas consists only of hydrogen, and that the energy is high enough (greater than ~ 10 GeV/amu) so ionization losses can be neglected, the flux f_i of a species i (where i is the atomic number) is simply related to the source term $Q_i(\mathrm{cm}^{-3}\mathrm{s}^{-1})$ and the mean escape length $\lambda_e(\mathrm{g\ cm}^{-2})$ through

$$\frac{f_i}{\lambda_{eff}} = \frac{Q_i}{mn_H} + \sum_j \delta_{i,j} \frac{f_j}{m} \tag{1}$$

$$\text{where } \frac{1}{\lambda_{eff}} = \frac{1}{\lambda_i} + \frac{1}{\lambda_e} + \frac{1}{\rho\beta C\tau_i}.$$

λ_i are the nuclear destruction lengths for species i due to interactions on interstellar material, τ_i is the decay lifetime for radioactive species ($=\infty$ for stable nuclei), ρ is the mean density in the storage column, and $\delta_{i,j}$ is the cross-section for producing nucleus i from nucleus j (λ_i decreases when i increases, e.g., $\lambda_{He} = 17$ g cm^{-2}, $\lambda_C = 7$ g cm^{-2}, $\lambda_{Fe} = 2.5$ g cm^{-2}).

For purely secondary species, such as the light elements lithium, beryllium, and boron, $Q_i = 0$ and the knowledge of the flux f_i and of the nuclear cross-sections involved is sufficient to determine the mean escape length λ_e; it is found to decrease as energy, or rigidity, increases: $\lambda_e \alpha R^{-0.6\pm0.1}$, as we will detail later. [Juliusson, Meyer, and Muller, 1972; Smith et al., 1973; Ormes and Protheroe, 1983; and Koch-Miramond et al., 1983]. (Rigidity is defined as the momentum per unit charge: $R = pc/eZ$).

As discrepancies are found between this simple picture and data, additional parameters are added to the phenomenological models to maintain agreement and improve understanding. One of the more widely used of these is the nested leaky-box model, really a two parameter leaky-box. In the original version of the nested leaky-box model [Cowsik and Wilson, 1973; Meneguzzi, 1973], cosmic rays are trapped both near their sources and at the boundaries of the galaxy, with a finite probability of escape from each. The assumption made by these authors is that λ_s, the pathlength traversed in the sources, but not that near the galactic boundary, is rigidity-dependent. The composition and the spectra of primaries and secondaries are essentially undistinguishable from those obtained with the energy-dependent leaky-box model, but in this case the galactic proton spectrum is identical to the injection spectrum, no matter what form $\lambda_s(R)$ has.

In the leaky-box model, the distribution of pathlengths around the mean is exponential. In contrast, the nested leaky-box model predicts a deficiency of short pathlengths. At high energy (E>1.5 GeV/amu), results of the HEAO3-C2 experiment, together with earlier results, can be accounted for with an exponential distribution of pathlengths [Protheroe, Ormes, and Comstock, 1981; Koch-Miramond et al., 1983]; however, lower energy data may require a truncation of the path length distribution [Garcia-Munoz et al., 1984].

b. Cosmic Ray Diffusion and Interstellar Turbulence Spectrum

It can be shown that the leaky-box model is equivalent to a diffusion model with a halo, provided the characteristic dimension of the storage volume is significantly larger than the galactic disk where the particles are presumably accelerated.

In most diffusion models, the elemental composition of cosmic rays is determined almost exclusively by one parameter, λ_e, related to the amount of matter traversed by the particles before escape; in general, λ_e is inversely proportional to the diffusion coefficient κ (in one-dimensional models, or in three-dimensional models with scalar diffusion) or to the component of the diffusion tensor perpendicular to the galactic plane. The constant of proportionality contains all the information on the distribution of the sources and on the boundaries of the containment region. For instance, let us consider one-dimensional models, where the cosmic ray sources are embedded in the gas disk of uniform density n_0 and of height h; cosmic rays of velocity v diffuse outward through a halo of height H >> h [Ginzburg, Khazan, and Ptuskin, 1980]. The diffusion coefficient κ is assumed (probably incorrectly) to be constant in space. Then κ is related to the mean escape length λ_e, calculated with the leaky-box Formula (1) by:

$$\kappa = (n_0 Hhvm)/\lambda_e. \tag{2}$$

In terms of diffusion models, variations of the elemental composition of cosmic rays could be interpreted as implying that either κ or the size of the confinement region varies with particle energy (rigidity).

The biggest uncertainty is what to assume for one size of the halo. Using H = 6 kpc and taking n = 0.5 atoms/cm^3 and λ_e = 7 g/cm^2 (the value at about 1 GeV/amu) gives a diffusion coefficient κ = 10^{28} cm^2/sec. Assuming that the particle transport is diffusive, what is responsible for the interactions which scatter the particles so effectively? Fermi [1949] has pointed out that moving inhomogeneities with a scale larger than the particles gyroradius in the magnetic field reflect particles of large pitch angle. This scattering process can lead to both diffusion and acceleration of cosmic rays. But the Fermi acceleration

mechanism has difficulties in satisfying the energy requirements and in explaining the observed abundances of secondary nuclei. In the last ten to fifteen years, the work on cosmic ray propagation has mostly concentrated on another process: resonant scattering of cosmic rays by hydromagnetic waves whose scales are comparable to their radius of gyration [Wentzel, 1974, and references therein]. This scattering leads to cosmic ray diffusion along the magnetic field lines; there is some energy exchange between cosmic rays and the hydromagnetic waves, but only to higher order in v_A/c, where $v_A = (B^2 4\pi\rho^*)^{1/2}$ ~ is the Alfvén velocity, where ρ^* is the density of ionized matter. The Alfvén velocity is in the range of tens of km/sec.

Let us define F(k) as the energy density in hydromagnetic waves per logarithmic bandwidth d(log k), relative to the ambient magnetic energy density $(B^2/8\pi)$. Then, in the framework of the quasi-linear theory (applicable if $F \ll 1$), the diffusion coefficient along field lines of particles of rigidity R and velocity v is given by:

$$\kappa\,(R) = \frac{4}{3\pi}\,\frac{v\,R/Bc}{F(k = r_e^{-1})}\,,\ \text{where}\ r_e = \frac{R}{Bc}\,. \tag{3}$$

The spectrum of hydromagnetic turbulence F(k) in the interstellar medium is extremely difficult to determine. Various methods exist that can lead to estimates or upper limits of the density spectrum of irregularities in the distribution of thermal electrons. Presently available results have been compiled by Armstrong, Cordes, and Rickett, [1981]. These authors conclude that the data are consistent with a power law spectrum of fluctuations, with an index of -3.6 ± 0.2. If the hydromagnetic wave spectrum had the same slope, this would be equivalent to:

$$F(k)\ \alpha\ k^{-0.6\pm0.2}. \tag{4}$$

A spectrum of this type may be the result of a cascade of turbulent energy in the interstellar medium from long scales to successively shorter scales; the turbulence at long scales is fed by cloud motions, which in turn are regenerated by supernova explosions. Kraichnan [1965] has argued that a cascade in an incompressible, weakly turbulent magnetized fluid, leads to a spectrum F(k)

$\alpha\ k^{-0.5}$. Such a cascade is energetically feasible in the hot phase ($T \sim 10^6$ K, $n \sim 10^{-3}\ cm^{-3}$) of the interstellar medium. If $F\ \alpha\ k^{-0.5}$ there, then we see from the Formula (3) that the cosmic ray diffusion coefficient is $\kappa\ \alpha\ vR^{0.5}$; this is very close to the dependence required to account for the observed variations of the ratio of secondary to primary nuclei with energy. Thus, the present observations of elements heavier than He, at energies lower than ~ 1000 GeV/amu, are well accounted for by a model where cosmic rays are scattered by resonant hydromagnetic waves related to the general interstellar turbulence [Cesarsky, 1975, 1980].

All the models discussed in this section assume that the only energy changes that cosmic rays undergo between production and detection are ionization losses in the interstellar medium and adiabatic losses during solar modulation. If cosmic rays are accelerated (or decelerated) by some additional mechanism while propagating, secondary particles get transferred to higher (lower) energies, and the secondary/primary profile as a function of energy is altered [e.g., Fransson and Epstein, 1980; Silberberg et al., 1983; and Simon, Heinrich, and Mathis, 1986]. The fact that the data at rigidities above a few GV are well explained by a variety of models indicates that new discriminators must be found to determine whether re-acceleration or deceleration is an important effect.

3. OBSERVATIONS OF COSMIC RAYS

With this theoretical picture in mind, let us turn to the observations made at Earth on the cosmic rays themselves.

Calorimetric and emulsion chamber devices have now measured [Grigorov, et al., 1971; Ryan, Ormes, and Balasubrahmanyan, 1972; and Burnett et al., 1983] the proton spectrum up to 100 TeV directly, and find that it obeys a power law $dN/dE = kE^{-\gamma}$ with $\gamma = 2.7 \pm 0.1$. There is no evidence of, but rather poor limits on, possible structure in the form of bumps, wiggles, or bends in the spectrum. This data is summarized in Figure 3. The proton differential spectrum does not appear to suffer any drastic change of slope between 10 and 10^6 GeV. The significance of these proton observations—the most abundant species of cosmic ray—is that the lack of structure implies

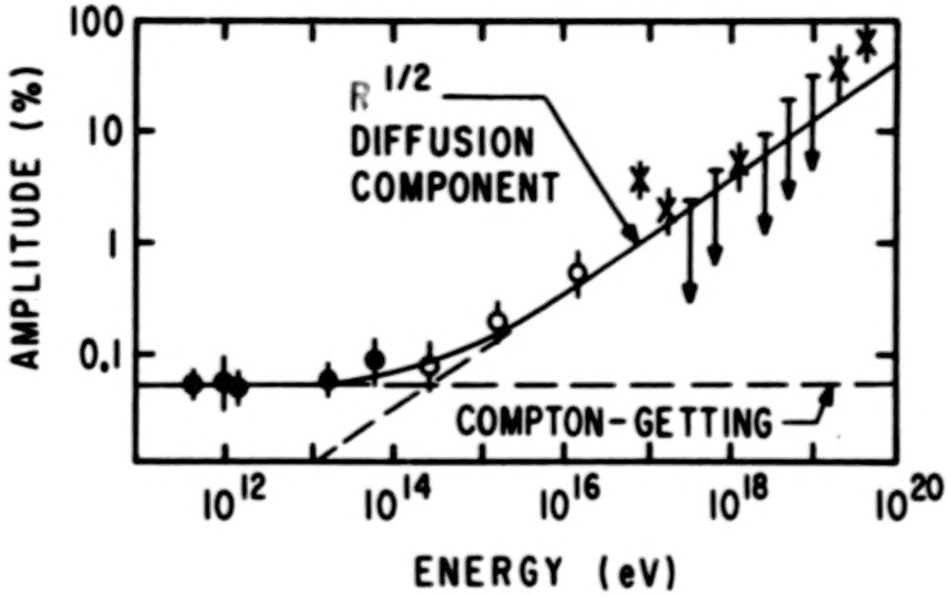


Figure 3. Direct measurements of cosmic ray spectra between 10 – 10^6 GeV as a function of total energy per nucleus. Measurements of the primary proton and helium spectrum are shown. The total particle spectrum is also shown [from Webber, 1983].

the mechanism(s) responsible for determining the shape of the proton spectrum is (are) continuous over this very large energy range.

In the leaky-box model, the mean confinement time of particles, τ_e, is proportional to λ_e. Neglecting nuclear losses, the cosmic ray density f is related to the source term Q through $f_i = Q\tau_e$. Under the plausible assumption that the source spectrum is a power law, $\tau_e(R)$ must also be a power law at least up to ~ 10^6 GV. This is a severe constraint on acceleration and propagation models as these two processes are presumably responsible for determining this spectral shape.

Several balloon measurements of cosmic ray composition at energies up to 150 GeV/amu have shown that the ratio of secondary to primary abundances

decreases as the energy increases [Juliusson, Meyer, and Muller, 1972; Ormes and Freier, 1978, and references therein]; also, the observed spectra of heavy primary species are flatter than those of lighter ones. More recently, the French-Danish spectrometer (C2) on the satellite HEAO-3 has provided extremely accurate data on the cosmic ray elemental composition from boron to zinc in the energy range 0.8-25 GeV/amu [Koch-Miramond, 1981].

Using these data, Formula (1) makes it possible to calculate λ_e as a function of R, at least in the context of the leaky-box model. Koch-Miramond et al. [1983] have corrected the low energy part of their data for the effects of solar modulation, assuming a modulation parameter ϕ = 600 MV, which is appropriate for the time in the solar cycle at which the measurements were made. They find that, at R>5.5 GV, and escape length $\lambda_e = 22\ R^{-0.6}$g/cm^2 of pure hydrogen accounts for the secondaries of C, O, and Fe. Ormes and Protheroe [1983] obtained a similar result. These analyses are limited by knowledge of cross-sections rather than by statistical uncertainties.

Many of the cross-sections have now been (or are currently being) measured and the resulting escape length is shown in Figure 4. Note its decrease with increasing energy, implying that the higher the energy, the more easily particles can escape the storage region. We now have confirmation that this decrease continues beyond a hundred GeV/amu. The data from the HEAO-C ultraheavy experiment were presented recently [Jones et al., 1985]. This experiment contained a complement of large area detectors designed to identify trans-iron nuclei. It had excellent statistics and could study nuclei heavier than calcium. Taking advantage of the relativistic rise of signals in the ionization chambers of their instrument, they obtained results on the abundances of several elements from 10 GeV/amu up to an energy > 100 GeV/amu. Their results are consistent with those of the French-Danish group in the range (10-25 GeV/amu) where both experiments apply; at higher energies, the HEAO3-C3 data indicate that the power law dependence with energy of the ratios (iron secondaries/iron) derived by the HEAO3-C2 data extends to about 100 GeV/amu, or rigidities of about 200 GV.

Streitmatter et al. [1985] reported that the iron spectrum itself has a slope of 2.65 in the energy range beyond 50 GeV/amu as expected in the leaky-box

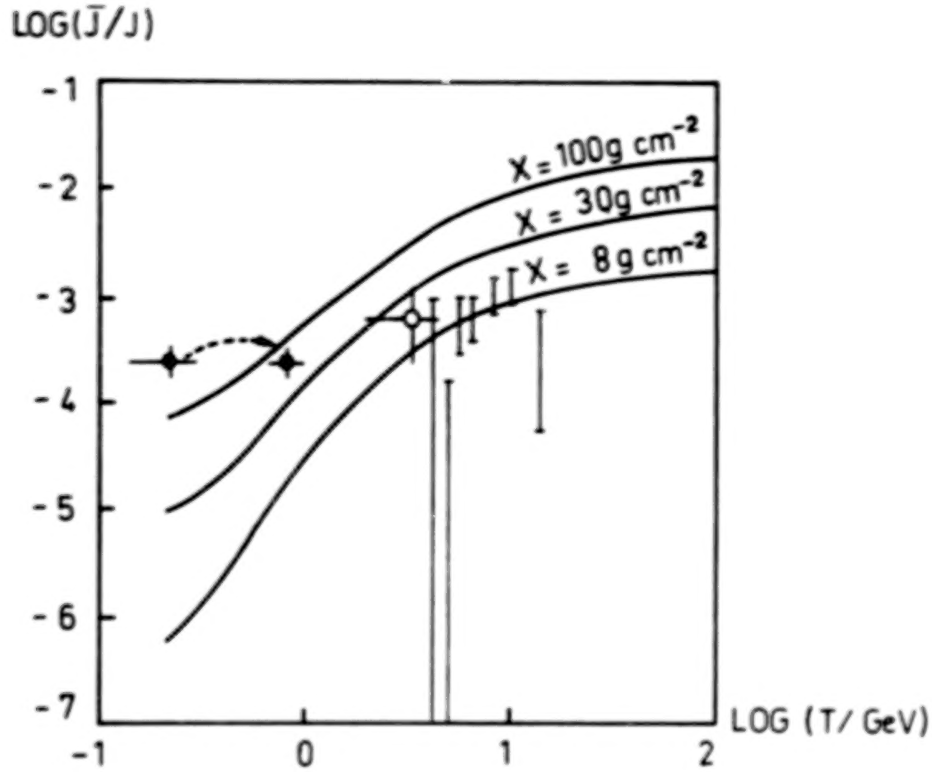


Figure 4. The mean escape length as a function of rigidity for a modulation parameter $\phi = 600$ MV [from Koch-Miramond et al., 1983].

model. What happens at even higher energies, or rigidities, such as R > 1000 GV? The ratio of secondary to primary element abundances have not yet been measured at such rigidities. Soon, results from the flight of the University of Chicago's Spacelab 2 experiment [L'Heureux et al., 1985] should solidify and extend these results. At still higher energies, in the decades on either side of 10^{15} eV or 10^{6} GV, the only method for learning about the spectrum and composition is through ground-based air shower studies. Hillas [1981] reviewed the situation of these energies a few years ago, and Linsley [1983] reviewed it more recently. Alternative points of view have been discussed in recent papers

by Balasubrahmanyan et al. [1987] and Streitmatter et al. [1985]. The situation is confusing. Some experiments indicate a gradual enrichment in the abundances of heavy nuclei, others do not. Some experiments indicate a hardening in the all particle spectrum above 10^{14} eV, others do not. New direct measurements in this energy range are surely needed.

a. Radioactive Secondary Nuclei

Measurements of the abundances of unstable secondary nuclei, such as ^{10}Be (with a mean lifetime for decay at rest of $\tau_d = 2.2 \times 10^6$ yr), ^{26}Al($\tau_d = 0.85 \times 10^6$ yr), and ^{35}Cl ($\tau_d = 0.45 \times 10^6$ yr), can bring some information on the mean age of cosmic rays, and/or help to determine the mean density in the storage volume, thus characterizing the different models.

In the framework of the leaky-box model, such measurements, combined with the determination of λ_e from the elemental composition, permit us in principle to estimate the mean escape time of cosmic rays and hence the mean gas density in the box. However, because most measurements are done at low energies, solar modulation again complicates the interpretation of the data.

Assuming that λ_e is energy independent, and using their own estimates of solar modulation effects, Wiedenbeck and Greiner [1980] deduce from their satellite data on ^{10}Be at 60-185 MeV/amu a confinement time of 8.4 (+4.0, −2.4) Myr, and a mean density $n_H = 0.33\ (+0.13, -0.11)\ cm^{-3}$. The mean age from ^{26}Al is 9 (+20, −6.5) Myr [Wiedenbeck, 1983], and ^{36}Cl leads to a lower limit to this age of 1 Myr [Wiedenbeck, 1985]. Since, in the solar neighborhood, the interstellar density (averaged over ~ 1 kpc in the disk) is estimated at $1-2\ cm^{-3}$, these results are generally interpreted as implying that galactic cosmic rays circulate in a low density halo which is at least 3 times thicker than the disk. However, they could also indicate that cosmic rays are preferentially trapped in low density regions of the disk between the clouds.

In diffusion models with a halo, radioactive isotopes formed in the disk often decay while passing through the halo. In that case, the average confinement time of particles in the galaxy may be much larger than the observed "mean

age" [Ginzburg, Khazan, and Ptuskin, 1980]. For instance, in the one-dimensional model described earlier, the abundances of secondary radioactive elements of decay period τ_d are determined by two combinations of parameters: $(n_0\tau_d)$ and (H/h). In principle, observations of the energy dependence of the abundance of isotopes of mean life at rest of ~ 10^6 years, at energies > 1 GeV/amu, should help constrain these parameters [e.g., Cesarsky et al., 1981].

b. Electron Spectrum

According to several recent measurements, the electron spectrum is parallel to the proton spectrum in the energy range 2-10 GeV; in this range, the electron flux amounts to ~ 1% of the proton flux. At higher energies > 50 GeV, the spectrum steepens, but electrons are still present at least up to 2000 GeV [Tang, 1984; Nishimura et al., 1980; Prince, 1979 and references therein]. A steepening of the high energy spectrum is expected, since the lifetime of a 30 GeV electron against radiation losses in the interstellar medium is ~ 10^7 years.

The observed electron spectrum does not impose strong constraints on the models proposed to explain the cosmic ray composition. It is important to remember that the equations describing the behavior and the energy changes of high energy electrons diffusing through the interstellar medium cannot be approximated by results obtained using the leaky-box model. In diffusion models, the distribution of the sources plays an important or even a predominant role. In addition, the injection spectrum of electrons is not known and can generally be adjusted to ensure that a given model fits the data.

4. ORIGIN AND PROPAGATION OF DIFFERENT COSMIC RAY SPECIES

a. Source Spectral Index, Composition and Energetics

After so many years of active research, there is not yet a firm answer to the question: where do cosmic rays come from? The main problem is, of course,

that the arrival direction brings little or no information on the source. Astrophysicists are then left with a less direct set of clues: spectrum, composition, energetics, anisotropy.

Observations must first be corrected for propagation effects; this is usually done in the framework of the galactic leaky-box model. Once λ_e is derived, at various rigidities, by applying Formula (1) (or its equivalent, including ionization losses, at energies < a few GeV/amu) to secondary species, it becomes possible to derive the source abundances Q_i by applying the same formula to primary species. In this way, Engelmann et al. [1985] have derived source spectra of primary species with $Z > 5$ from the HEAO3-C2 data. Assuming H and He nuclei behave like the other species, the observed spectrum must be divided by $\lambda_e(R)$ in order to correct for propagation. The source spectra thus obtained are displayed in Figure 5. Data from other experiments are also represented. In the range $R \simeq 2 - 20$ GV, Engelmann et al. [1985] found that the spectra are generally steeper than previously thought. This leads to a rather surprising conclusion that the source spectra of the heavy nuclei are steeper (index 2.4) than those of the more abundant protons (index 2.1). Unfortunately, this is based on an experiment which only covers a narrow band of energy—the lower end of which may be complicated by solar modulation effects—and the experiment itself may be subject to systematic effects at the high energy end. Therefore, the results need confirmation. However, they are suggestive that there may be more than one "source" or mechanism operating to produce the locally observed cosmic rays.

The implications of this result have not yet been studied in full detail. Essentially all of the published work on cosmic ray origin continues to assume that protons and alpha particles originate and propagate as the other species, and that the λ_e derived from studies of heavy nuclei can be used to estimate the energetics. For the local Kpc^2 in the galactic plane, cosmic ray energetics is derived using the fact that, on the average, cosmic rays escape at a rate $c\lambda_{gal}/\lambda_e$, where λ_{gal} is the column density of matter across the galactic disk. The energy requirement to maintain the cosmic ray pool is then $\sim 10^{38}$ erg/Kpc^2 sec. (Alternative derivations, using the cosmic ray "age" derived from secondary radioactive isotopes, yield similar results). If we retain the same leaky-box model for all species, the results of Engelmann et al. [1985]

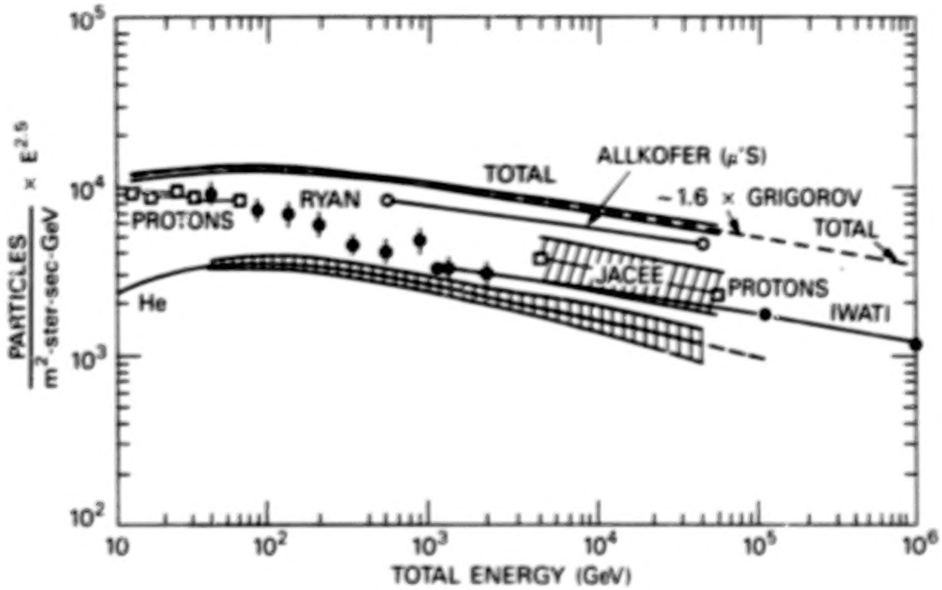


Figure 5. Observed spectra compared with various source spectra. Proton and possibly helium are consistent with source spectra with indices 2.1. Heavier nuclei, on the other hand, may require steeper source spectra. [This figure is from Engelmann et al., 1985. Original references can be found therein.]

imply that the local cosmic rays consist of two components: a flat component, with source index ~ 2.1, and a steep component, with source index $\simeq$ 2.4. At rigidities below ~ 100 GV, most of the nuclei heavier than He would belong to the steep component, while at all energies the flat component would be dominant in the proton flux.

The leaky-box formalism, as we have seen, accounts well for the observations relating to the steep component which is rich in heavy nuclei. But there is no compelling reason to believe that the flat component, which is relatively proton-rich, has the same history. The steep component may be just local, and transient; the determinations of λ_e and of age from radioactive isotopes only relate to this component. But the proton-rich component is the only one that counts when discussing energetics, constancy in time of the cosmic ray

flux, and isotropy. The abundances of secondary elements with $Z > 2$, at energies < 100 GeV, may simply not be relevant when studying it!

Some light can be focused on this problem by refining the spectra of hydrogen and helium, and studying carefully their secondaries ^3He, D, and antiprotons.

b. Antiprotons

The general picture of cosmic ray storage and propagation in the galactic magnetic fields described above has been based largely on the abundances of heavy nuclei. However, recent observation of the antiprotons has thrown this unified picture into disarray.

Secondary antiprotons are generated in the inelastic collisions between high energy nuclear cosmic rays and interstellar medium particles. The flux of galactic antiprotons has been measured recently by Golden et al. [1979 and 1984], by Bogomolov et al. [1979] and by Buffington, Schindler, and Pennypacker [1981] at various energies (Figure 6). The data of Golden et al. seem to be on solid ground and to be confirmed by the lower statistics observation of Bogomolov et al. The low energy point of Buffington et al. is more startling and unfortunately on less stable ground experimentally [Stephens, 1981].

Buffington, Schindler, and Pennypacker [1981] measured the flux of cosmic ray antiprotons in the range 130-320 MeV, which corresponds after demodulation to a mean interstellar energy of ~ 800 MeV. The data are compared with the calculation based on the leaky-box model where antiprotons are assumed to be secondaries produced by collisions of protons and heavier nuclei with interstellar matter. The Buffington et al. point falls well below the kinematic cutoff but indicates that there is a high flux of low energy antiprotons present [Buffington and Schindler, 1981]. Even ignoring this data, the mean target thickness to produce the intensity observed by Golden et al. must be three or four times that of the heavier cosmic rays. This is discussed at length in the paper by Balasubrahmanyan, Ormes, and Streitmatter (this volume) where the various models which have been advanced as an explanation have been presented. Combined with the finding that the source spectra of heavier

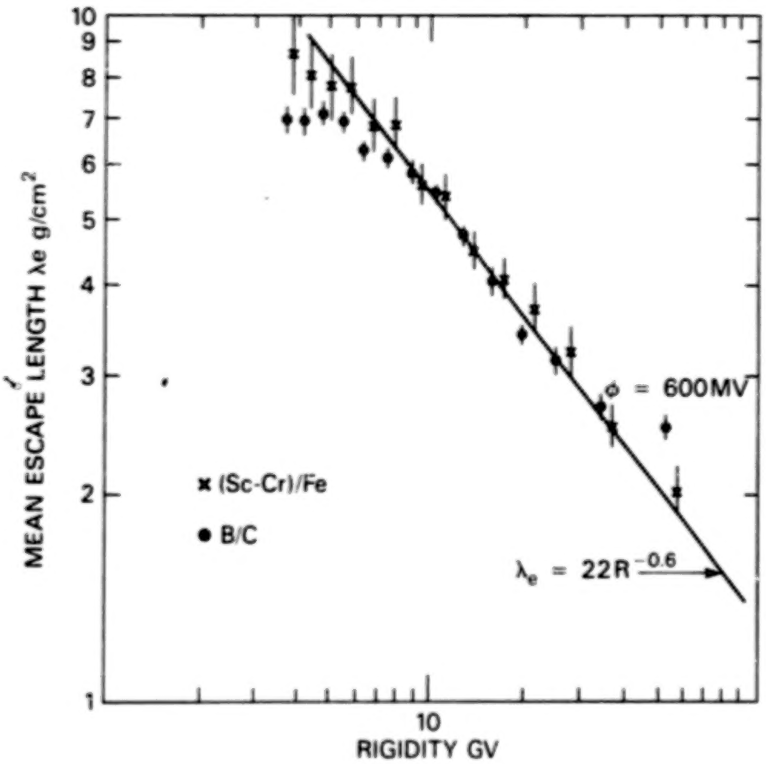


Figure 6. The observed antiproton to proton flux ratios [Golden et al., 1984, vertical bars; Bogomolov et al., 1979, open circle; and Buffington, Schindler, and Pennypacker et al., 1981, solid circle] compared with the antiprotons produced by a shell of matter surrounding a s: ong shock acceleration region [from Lagage and Cesarsky, 1985].

nuclei may be different from protons and helium, these data may indicate that the origin of the protons and/or their history after acceleration is different from that of heavier nuclei.

It may be that this unexpectedly high abundance of antiprotons is an additional indication that the history of all cosmic rays does not follow from the abundances of secondary nuclei alone. Golden et. al. [1984], showed that within

the framework of an energy independent leaky-box model (source spectral index = 2.6), $21 g/cm^2$ of material is required. Alternately, Lagage and Cesarsky [1985] showed that the high energy observations of antiprotons could be accounted for if all cosmic ray protons had a source spectrum of index 2.1 and traversed $7 g/cm^2$ in their sources before escaping into the galaxy, or if a fraction x of the cosmic rays traversed a slab of width X at the source, with $xX = 7 g/cm^2$ [Lagage and Cesarsky, 1985]. (This $7 g/cm^2$ is energy-independent and should not be confused with the $7 g/cm^2$ traversed by heavy nuclei at 1 GeV/nucleon.) As noted by these authors, a problem with this "thick-source" model is that, in addition to the antiprotons, neutral pions are produced, which decay into gamma rays. The total galactic gamma ray flux predicted by this model exceeds that observed by COS-B by a factor ~ 3.

If protons have a different history from heavier nuclei, what about helium nuclei? There are data from a balloon experiment indicating that at high energy the helium nuclei may have traversed a target intermediate between that of protons ($21 g/cm^2$) and heavier nuclei ($7 g/cm^2$). This result [Jordan and Meyer, 1984] is sensitive to the assumed shape of the helium spectrum and remains controversial. Further data on the ^3He and deuterium abundances at high energy are needed to resolve this issue: is the matter traversed a continuous function of atomic number, or is there a discrete difference between protons and all heavier nuclei? If the latter, to which camp do the helium abundances belong?

c. Anisotropy

We have been taking the point of view that abundances of the elements are indicative of cosmic ray propagation. An alternative point of view has been taken by Hillas [1984], who uses the anisotropy as the main indicator on the propagation. This can only be done at energies above a few 100 GeV, since at lower energies the trajectories of the cosmic rays are perturbed by the solar wind. Hillas notes that, at energies > 10^3 GeV, the amplitude of the first harmonic of the cosmic ray anisotropy is, very roughly proportional to the product (cosmic ray differential flux . $E^{2.47}$) (reference Figure 1). Now, if τ_e is the confinement time, the anisotropy is expected to be ~ t/τ_e, where t is the time for escape in a straight line. Hillas proposes a simple interpretation of Figure 1: that the source spectrum is a power law of index 2.47 over the whole energy

range, and that all the features in the spectrum are due to propagation effects. At 10^3 GeV, the amplitude of intensity variation is of $\sim 0.06\%$. If the boundary of the cosmic ray confinement region is at y kpcs, $\tau_e(10^3 \text{ GeV}) \simeq$ 5y Myr. Since the spectrum of protons does not appear to change significantly between 5 and 10^3 GeV (see Figure 3), the mean age at 5 GeV would be $\sim$ $(1000/5)^{2.7-2.47}\ \tau_e\ (10^3 \text{ GeV}) \simeq 17$ Myr (where 2.7 is the observed index of the proton spectrum at these energies). This is comparable to the age derived from radioactive secondary isotopes, so the global energetics of galactic cosmic rays is not very much changed in this picture.

d. Cosmic Ray Sources

We summarize by listing the requirements on cosmic ray sources.

i) Energetics: the order of magnitude of the power required to replenish cosmic rays "within" a cylinder of base 1 Kpc^2 within the galactic disk, of height 1 to $\sim$ several Kpc, is $\sim 10^{38}$ ergs/sec.

ii) Source spectrum: most probably a power law, at least in the range from a few GeV/amu to $\sim 10^6$ GeV/amu, perhaps up to 10^8 or even 10^9 GeV! Spectral index: 2.1? 2.4? or 2.7? Or somewhere in this range.

iii) Source composition: well determined now, for most elements, in the GeV/amu range. May give clues to the origin of the cosmic radiation or at least, as we have seen, to a component of it.

Within a radius of 3 kpc from the Sun, the average energy input from supernovae is estimated to be $\sim 10^{39}$ erg s^{-1} kpc^{-2}; supernovae are widely believed to be the main accelerators of cosmic rays. Stellar winds expend $\sim$ 10^{38} erg s^{-1} kpc^{-2} in the interstellar medium, and they may also contribute to cosmic ray acceleration. [Cesarsky and Montmerle, 1983]. Composition arguments have often been invoked to eliminate pulsars as a candidate source, but the debate on the role of pulsars in cosmic ray acceleration is not closed.

5. ACCELERATION MECHANISM

We require (probably) an acceleration mechanism capable of producing a power law spectrum.

a. Fermi Acceleration

The basic concept of acceleration of particles via encounters with "moving magnetic walls" was introduced by Fermi as early as 1949. Fast particles of velocity v that encounter magnetic walls are separated by a mean distance moving at a velocity V. The walls reflect the particles and enhance their energies.

$$\frac{dE}{dt} = \frac{2V^2}{c\lambda} E = \alpha E \tag{4}$$

This process has enjoyed an enduring popularity among astrophysicists because it predicts that the energy spectrum of the colliding particles should be a power law: $N(E) \propto E^{-\gamma}$, $\gamma = (1 - \frac{1}{\alpha\tau_e})$, where τ_e is the mean time spent by a particle in the accelerating region.

b. Particle Acceleration by Parallel Shocks in a Scattering Medium

This attractive mechanism must have been in the air several years ago, as it has been discovered simultaneously by astrophysicists all over the world [Krimsky, 1977; Axford, Leer, and Skadron, 1977; Blandford and Ostriker, 1978; and Bell, 1978]. This is somewhat surprising, as the tools used in the various derivations, and the motivation, have been around for a much longer time.

Let us consider a strong shock, propagating at a velocity V in the direction of the magnetic field lines. We assume that $V = v_A$, where v_A is the Alfvén velocity. In the shock frame, the gas is flowing in at a velocity $u_1 = V$. At the shock, the gas is compressed by a factor r, so that the velocity downstream, relative to the shock, is $u_2 = V/r$.

The presence of scattering centers of cosmic rays is postulated, so that cosmic rays diffuse on both sides of the shock; the diffusion coefficient is, in general,

a function of space, particle momentum, and time. In any case, the scattering centers act as cosmic ray traps, ensuring that the particles will be reflected back and forth across the shock a large number of times. Every passage through the shock is equivalent to running head-on into a "magnetic wall" of velocity $V = u_1 - u_2 = V(1 - 1/r)$; averaged over all incidence angles, there is a mean energy gain per traversal of the shock given by

$$\Delta E = (4/3)\,(V/c)\,(1 - 1/r)E \tag{5}$$

Taking proper account of the probability of particles escaping the system leads to the time-independent spectrum:

$$N(E) \propto E^{-\mu},\ \mu = (2+r)/(r-1) \tag{6}$$

For strong adiabatic shocks, $r = 4$ and $u = 2$. Weaker shocks generate steeper spectra.

The remarkable property of this mechanism is that, in the time-independent limit, the slope of the power law it generates depends only on the shock strength, and not at all on the diffusion coefficient (assumed "small enough") or the dimensions of the scattering region (assumed "large enough").

The study of shock acceleration of cosmic rays is now an active area of research. A fundamental review of the subject has been written by Drury [1983]. A detailed application of the mechanism to the acceleration of galactic cosmic rays is given in Blandford and Ostriker [1980]; see also Axford [1981].

Many aspects of this mechanism have been studied since, and it is impossible to review this rich field here. Let us just emphasize some of the main problem areas:

i) This problem has always been treated in the framework of the quasi-linear theory, which assumes that the turbulent energy in the hydrodynamic waves acting as particle scatterers is much less than the energy density of the magnetic field. However, the anisotropies induced by supernova shocks in the pre-existing population of galactic cosmic rays are sufficient to render

these waves extremely unstable; the wave amplitudes predicted by the qausi-linear theory are too high to be fully consistent with this theory.

ii) If cosmic rays extract so much energy from the shock, their pressure can become the dominant one. For instance, this will inevitably occur if cosmic rays are getting accelerated by a strong shock, to a spectrum E^{-2}, for a sufficiently long time. Even if the shock is not so strong ($r < 4$), the cosmic ray pressure can become dominant if the rate of injection of particles in the system is sufficiently rapid. The expectation is that, eventually, the cosmic rays broaden the shock, making it a less efficient particle accelerator. If the shock becomes wider than the particle mean free path λ, all particles of a given energy obtain the same amount of adiabatic acceleration as they cross the shock region. Ellison and Eichler [1985] have studied these problems and find that this mechanism still produces a universal spectrum which is very similar to a power law of index ~ 2. The efficiency of cosmic ray acceleration by this mechanism is very high, of order 25%.

iii) An important problem of the theories of shock wave acceleration is that the maximum energy that can be attained is limited, either by the lifetime of the shock itself or by its curvature radius. This problem was treated in detail by Lagage and Cesarsky [1983, 1985]. In the case of supernova shocks, the limiting factor is the shock lifetime; under most optimistic assumptions, the maximum energy E_{max}, for particles of charge Z, is only $\sim 10^5$ $Z(B/10^{-6}$ gauss) GeV, where B is the strength of the magnetic field in the most diffuse phase of the interstellar medium.

This result holds whether the shock is linear or cosmic ray dominated. Taking into account the nonlinearity introduced by the fact that, upstream, the Alfvén waves are generated by the cosmic rays, so that the diffusion coefficient is space- and time-dependent, E_{max} is limited to values which may be as low as 2000 Z $(B/10^{-6}$ gauss) GeV. Invoking supernova shocks propagating in the galactic halo does not alleviate the problem [Lagage and Cesarsky, 1987].

The possible acceleration of high energy cosmic rays by stellar wind terminal shocks is still controversial. If shock acceleration is operating there over long times, stellar winds have the advantage that the shock is a standing shock,

which remains strong for longer times than supernova shocks. The maximum energy is then determined by the shock curvature, and the strength of the magnetic field: $E_{max} \sim 5.10^5 Z(B/10^{-5}G)$ (D/5 pc) GeV, where D is the shock radius. In a recent paper Kazanas and Ellison [1986] attempt to model the binary X-ray source Cygnus X-3, which may be emitting ultrahigh energy gamma rays [Samorski and Stamm, 1983; Lloyd-Evans et al., 1983; Watson, 1985 and references therein], and thus be a source of cosmic rays of energy up to 10^7-10^8 GeV. Assuming the presence of a collisionless, spherical accretion shock around the compact object in Cygnus X-3, and assuming that the magnetic field strength is in equipartition with the accretion flow, Kazanas and Ellison argue that protons of energy as high as 7.10^6 GeV may be accelerated by the shock in this system.

While these problems are serious and are being worked theoretically, it is clear that shocks in the interstellar medium do accelerate particles. Shock acceleration remains the most promising mechanism for producing the power law spectra observed in the galactic cosmic rays, at least in the energy range from 1 to 10^6 GeV.

6. SUMMARY

The study of systematic trends in elemental abundances is important for unfolding the nuclear and/or atomic effects that should govern the shaping of source abundances and in constraining the parameters of cosmic ray acceleration models [for reviews see Casse, 1984; Simpson, 1983]. These issues were discussed in the rapporteur paper by J. P. Meyer [1985]. The isotopic composition and elemental abundances of trans-iron nuclei have much to contribute about the nucleosynthesis, sites, and timescales for the origin of cosmic rays. [See Binns et al., 1987, this volume.]

In principle, we can also learn much about the large-scale distributions of cosmic rays in the galaxy from all-sky gamma ray surveys such as COS-B and SAS-2. Gamma ray intensities are proportional to the line integral along the line of sight of the product of the cosmic ray flux and the matter density. However, because of the uncertainties in the matter distribution which come from the inability to measure the abundance of molecular hydrogen, the results

are somewhat controversial. A debate exists as to whether on a scale of 0.5 to 1 kpc there are more cosmic rays where there is more matter. [See paper by Fichtel, 1987, this volume.] Questions exist about whether the cosmic ray intensity falls off in the outer galaxy or remains about the same. Because around 100 MeV there are almost as many gamma rays from the bremsstrahlung process as from π^0 decay, resolution of these issues will await the improved energetic gamma ray experiment telescope (EGRET) on GRO. Very high energy ground-based cosmic ray telescopes will help in understanding the role that sources like Cygnus X-3 play in accelerating cosmic rays. High resolution radio observations of external galaxies [e.g., Duric et al., 1986] may provide clues about the role of shocks and spiral density waves in particle acceleration.

As we have seen, the leaky-box model accounts for a surprising amount of the data on heavy nuclei. However, a growing body of data indicates that this simple picture may have to be abandoned in favor of more complex models which contain additional parameters. For example, an energy-dependent modification of the exponential path length distribution, natural to the simple leaky-box model, has long been invoked to explain differences between the escape length derived from sub-iron secondaries and the Li, Be, and B components [Guzik et al., 1985]. The shape of the high energy electron spectrum led Tang and Muller [1983] to favor the nested leaky-box model. The spectral differences at the source and the antiproton observations lead us to postulate a separate origin for protons and heavier nuclei. Acceleration by weak shocks may lead to a reinterpretation of the observed element ratios in terms of material traversed. Observations of anisotropy, the consistency of the flux in time, and the gamma ray distribution tell us primarily about properties of protons, and nothing about those same quantities for heavier nuclei. Age measurements have been made for low energy nuclei and possibly for electrons, but their interpretations are model-dependent. In short, the observations still leave us in some confusion and greatly in need of further observations.

Future experiments on the Spacelab and Space Station will hopefully be made of the spectra of individual nuclei at high energy. Antiprotons must be studied in the background free environment above the atmosphere with much higher

reliability and precision (the world's observed antiprotons number of order 60) to obtain spectral information.

Isotopic composition needs to be measured over more elements and over an extended energy range. Ultraheavy abundances beyond tin in the periodic table must be measured with single element resolution and the abundances of actinides determined.

The future for these observations includes the Heavy Nuclei Collector currently being constructed for an exposure on NASA's Long Duration Exposure Facility and the Particle Astrophysics Superconducting Magnet Facility (Astromag) being planned for NASA's Space Station. The Gamma Ray Observatory is scheduled for launch in 1990. If all these plans are brought to fruition, the next two decades should see tremendous progress made in unraveling the problem of the origin of cosmic rays.

REFERENCES

Allen, R. J., Baldwin, J. E., and Sancisi, R., 1978, *Astron. and Astrophys.*, **62**, 397.

Armstrong, J. W., Cordes, J. M., and Rickett, B. J., 1981, *Nature*, **291**, 561.

Axford, W. I., 1981, *17th Internat. Cosmic Ray Conference Papers (Paris)*, **12**, 155.

Axford, W. I., Leer, E., and Skadron, G., 1977, *15th Internat. Cosmic Ray Conference Papers (Plovdiv)*, **11**, 131.

Balasubrahmanyan, V. K., Ormes, J. F., and Streitmatter, R., 1986, this volume.

Balasubrahmanyan, V. K., Sreekantan, B. V., Goodman, J. A., and Yodh, G. B., 1987, to be published.

Bell, A. R., 1978, *MNRAS*, **1982**, 147.

Binns, W. R. et al., 1987, this volume.

Blandford, R. D., and Ostriker, J. P., 1978, *Astrophys. J.*, **221**, L129.

Blandford, R. D., and Ostriker, J. P., 1980, *Astrophys. J.*, **237**, 793.

Bogomolov, E. A., Lubyanaya, N. D., Romanov, V. A., Stephanov, S. V., and Shulakova, M. S., 1979, *16th Internat. Cosmic Ray Conference Papers (Kyoto)*, **1**, 330.

Bradt, H., and Peters, B., 1948, *Phys. Rev.*, **74**, 1828.

Buffington, A., and Schindler, S. M., 1981, *Astrophys. J.*, **247**, L105.

Buffington, A., Schindler, S. M., and Pennypacker, C. R., 1981, *Astrophys. J.*, **248**, 1179.

Burnett, T. H. et al., 1983, *Phys. Rev. Letters*, **51**, 1010.

Casse, M., 1984, in *Stellar Nucleosynthesis*, ed. C. Chiosi and A. Renzini (Dordrecht: D. Reidel Publishing Co.), p. 55.

Cesarsky, C. J., 1975, *14th Internat. Cosmic Ray Conference Papers*, **12**, 4166.

Cesarsky, C. J., 1980, *Ann. Rev. Astron. Astrophys.*, **18**, 289.

Cesarsky, C. J., Koch-Miramond, L., and Perron, C., 1981, *17th Internat. Cosmic Ray Conference Papers (Paris)*, **2**, 22.

Cesarsky, C. J., and Montmerle, T., 1983, *Sp. Sci. Rev.*, **36**, 173.

Cowsik, R., and Wilson, L. W., 1973, *13th Internat. Cosmic Ray Conference Papers (Denver)*, **1**, 500.

Drury, L. O. C., 1983, *Rep. Progress in Physics*, **46**, 973.

Duric, N., Seaquist, E. R., Crane, P. C., and Davis, L. E., 1986, *Astrophys. J.*, **304**, 82.

Ellison, D. C., and Eichler, D., 1985, *Phys. Rev. Letters*, **55**, 2735.

Engelmann, J. J. et al., 1985, *Astron. and Astrophysics*, **148**, 12.

Fermi, E., 1949, *Phys. Rev.*, **75**, 1169.

Fichtel, C., 1987, this volume.

Fransson, C., and Epstein, R. I., 1980, *Astrophys. J.*, **242**, 411.

Freier, P., Lofgren, E. J., Ney, E. P., and Oppenheimer, F., 1948, *Phys. Rev.*, **74**, 1818.

Garcia-Munoz, M., Guzik, T. G., Simpson, J. A., and Wefel, J. P., 1984, *Astrophys. J.*, **280**, L13.

Ginzburg, V. L., and Syrovatskii, S. I., 1964, *The Origin of Cosmic Rays* (New York: Pergamon Press).

Ginzburg, V. L., Khazan, Y. M., and Ptuskin, V. S., 1980, *Ap. Sp. Sci.*, **68**, 295.

Golden, R. L. et al., 1979, *Phys. Rev. Letters*, **43**, 1196.

Golden, R. L., Mauger, B. G., Nunn, S., and Horan, S., 1984, *Astrophys. Letters*, **24**, 75.

Grigorov, N. L. et al., 1971, *12th Internat. Cosmic Ray Conference Papers*, **5**, 1746.

Guzik, T. G. et al., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **2**, 76.

Hillas, A. M., 1981, *17th Internat. Cosmic Ray Conference Papers (Paris)*, **13**, 69.

Hillas, A. M., 1984, *Ann. Rev. Astron. Astrophys.*, **22**, 425.

Jones, M. D. et al., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **2**, 28.

Jordon, S. P., and Meyer, J. P., 1984, *Phys. Rev. Letters*, **53**, 505.

Juliusson, E., Meyer, P., and Muller, D., 1972, *Phys. Rev. Letters*, **29**, 445.

Kazanas, D., and Ellison, D., 1986, *Nature*, **319**, 380.

Koch-Miramond, L., 1981, *17th Internat. Cosmic Ray Conference Papers (Paris)*, **12**, 21.

Koch-Miramond, L., Engelmann, J. J., Goret, P., Juliusson, E., Masse, P., and Soutoul, A., 1983, *18th Internat. Cosmic Ray Conference Papers (Bangalore)*, **9**, 275.

Kraichnan, R. H., 1965, *Phys. Fluids*, **8**, 1385.

Krimsky, G. F., 1977, *Dok. Akac. Nauk. SSSR*, **234**, 1306.

Lagage, P. O., and Cesarsky, C. J., 1983, *Astron. Astrophys.*, **125**, 249.

Lagage, P. O., and Cesarsky, C. J., 1985, *Astron. Astrophys.*, **147**, 127.

Lagage, P. O., and Cesarsky, C. J., 1987, to be published.

L'Heureux, J., Meyer, P., Muller, D., and Swordy, S., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **3**, 276.

Linsley, J., 1983, *18th Internat. Cosmic Ray Conference Papers (Bangalore)*, **12**, 135.

Lloyd-Evans, J. et al., 1983, *Nature*, **305**, 784.

Meneguzzi, M., 1973, *13th Internat. Cosmic Ray Conference Papers (Denver)*, **1**, 378.

Meyer, J. P., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **9**, 141.

Muller, D., and Tang, J., 1981, *17th Internat. Cosmic Ray Conference Papers (Paris)*, **9**, 142.

Nishimura, J. et al., 1980, *Astrophys. J.*, **238**, 394.

Ormes, J. F., 1983, *18th Internat. Cosmic Ray Conference Papers (Bangalore)*, **2**, 187.

Ormes, J. F., and Protheroe, R. J., 1983, *Astrophys. J.*, **272**, 756.

Ormes, J. F., and Freier, P., 1978, *Astrophys. J.*, **222**, 471.

Prince, T. A., 1979, *Astrophys. J.*, **227**, 676.

Protheroe, R. J., Ormes, J. F., and Comstock, G. M., 1981, *Astrophys. J.*, **247**, 362.

Ryan, M. J., Ormes, J. F., and Balasubrahmanyan, V. K., 1972, *Phys. Rev. Letters*, **28**, 985.

Samorski, M., and Stamm, W., 1983, *Astrophys. J.*, **268**, 17.

Silberberg, R., Tsao, C. H., Letaw, J. R., and Shapiro, M. M., 1983, *Phys. Rev. Letters*, **51**, 1217.

Simon, M., Heinrich, W., and Mathis, K. D., 1986, *Astrophys. J.*, **300**, 32.

Simpson, J. A., 1983, *Ann. Rev. Nucl. Part. Sci.*, **33**, 323.

Smith, L. H., Buffington, A., Smoot, G. F., Alvarez, L. W., and Wahlig, M. A., 1973, *Astrophys. J.*, **180**, 987.

Stephens, S. A., 1981, *Nature*, **289**, 267.

Streitmatter, R. E., Balasubrahmanyan, V. K., Ormes, J. F., and Acharya, B. S., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **2**, 40.

Streitmatter, R. E., Balasubrahmanyan, V. K., Protheroe, R. J., and Ormes, J. F., 1985, *Astron. and Astrophysics*, **143**, 249.

Tang, J., and Muller, D., 1983, *18th Internat. Cosmic Ray Conference Papers (Bangalore)*, **9**, 250.

Tang, K. K., 1984, *Astrophys. J.*, **278**, 881.

Watson, A. A., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **9**, 111.

Webber, W., 1983, in *Composition and Origin of Cosmic Rays*, ed. M. M. Shapiro (Dordrecht: D. Reidel Publishing Co.), p. 25.

Wentzel, D. G., 1974, *Ann. Rev. Astron. Astrophys.*, **12**, 71.

Wiedenbeck, M. E., and Greiner, E. D., 1980, *Astrophys. J. Letters*, **239**, L139.

Wiedenbeck, M. E., 1983, in *Composition and Origin of Cosmic Rays*, ed. M. M. Shapiro (Dordrecht: D. Reidel Publishing Co.), p. 65.

Wiedenbeck, M. E., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **2**, 84.

PART III:

GAMMA RAY, X-RAY, AND INFRARED ASTRONOMIES

N87
24258
UNCLAS

11

HIGH ENERGY GAMMA RAY ASTRONOMY

Carl E. Fichtel
Laboratory for High Energy Astrophysics
Goddard Space Flight Center
Greenbelt, Maryland 20771

ABSTRACT

High energy gamma ray astronomy has evolved with the space age. Nonexistent twenty-five years ago, there is now a general sketch of the gamma ray sky which should develop into a detailed picture with the results expected to be forthcoming over the next decade. The galactic plane is the dominant feature of the gamma ray sky, the longitudinal and latitudinal distribution being generally correlated with galactic structural features including the spiral arms. Two molecular clouds have already been seen. Two of the three strongest gamma ray sources are pulsars. The Vela pulsar, PSR 0833-45, exhibits two pulses in the gamma ray region, as opposed to one in the radio region, neither of them in phase with the radio pulse. The other of the strongest gamma ray sources is that at L = 195, b = +5); it had no obvious counterpart at other wavelengths when it was found and correlation at other wavelengths is still uncertain. The highly variable X-ray source Cygnus X-3 was seen at one time, but not another in the 100 MeV region, and it has also been observed at very high energies ($>10^{11}$ eV). Beyond our galaxy, there is seen a diffuse radiation, whose origin remains uncertain, as well as at least one quasar, 3C 273. Looking to the future, the satellite opportunities for high energy gamma ray astronomy in the near term are the GAMMA-I planned to be launched in late 1987 and the Gamma Ray Observatory, scheduled for launch in 1990. The Gamma Ray Observatory will carry a total of four instruments covering the entire energy range from 3×10^4 eV to

3×10^{10} eV with over an order of magnitude increase in sensitivity relative to previous satellite instruments. On the ground, there is the possibility of much more sensitive measurements above 10^{11} eV than exist now. In the more distant future, the NASA Space Station should provide opportunities to fly quite large gamma ray instruments which might be refurbished or reconfigured in space.

1. INTRODUCTION

Gamma ray astronomy is truly a recent, space age addition to the field of astronomy. When Frank McDonald joined Goddard Space Flight Center in 1959, gamma ray astronomy consisted largely of theoretical papers describing its important scientific potential. Motivated in part by these articles, which were in some cases quantitatively rather optimistic about intensities, scientists began to develop instruments to detect celestial gamma rays first to be flown on high altitude balloons and later satellites. At the moment, the promised rich potential of gamma ray astronomy is beginning to be realized in the exploration of several aspects of astrophysics, and, if the missions planned for the next several years come into being, their results, in combination with those from other frequency ranges of astronomy, will almost certainly provide us with new concepts of the evolution and nature of the universe.

What the theorists realized approximately three decades ago was that, as the energies of the individual gamma ray photons suggest, gamma ray astronomy relates very directly to the most energetic processes in the universe ranging from the scale of individual particle acceleration and interaction through the formative processes in the galaxy and stellar explosions, to the largest ensembles imaginable. The nucleonic galactic cosmic rays reveal themselves through the high energy gamma rays emitted by the π^0 mesons which are formed in nuclear interactions between cosmic rays and interstellar matter throughout the galaxy. High energy cosmic ray electrons reveal themselves through interactions with matter and photons. The region around a black hole is predicted to emit characteristic gamma rays, and the death of a black hole should reveal itself with the emission of a very specific type of high energy gamma ray burst.

Matter-antimatter annihilation produces another characteristic type of gamma ray spectrum. Neutron star pulsars have already been seen in the gamma ray frequency, as has at least one quasar.

In addition to its value in relation to the physical processes of the universe to which it speaks, gamma ray astronomy has the very attractive feature that the universe is largely transparent to gamma rays. They can reach the solar system from the galactic center, distant parts of the universe, and dense regions near the centers of active galaxies—regions which cannot be viewed in the optical or low energy X-ray region, since, in contrast to optical photons which penetrate easily through the Earth's atmosphere, only the total amount of matter and not its form is relevant for gamma ray interactions. A specific illustration of the penetrating power of this radiation is the following. A high energy gamma ray passing through the diameter of the central plane of the galactic disk has about a one percent chance of interacting for a typical path. By contrast, an optical photon can only penetrate about one-tenth the distance from the galactic center to the Earth in the central plane of the disk. This remarkable window extends from a few times 10^7 eV, below which it begins to close slowly as the energy decreases so that as the X-ray region is reached the center of the galaxy is quite opaque, to 10^{15} eV, at which point there begins a one to two decade region in energy wherein gamma ray interactions with the blackbody radiation are important.

In spite of its importance, gamma ray astronomy is the last major wavelength range to yield its wealth of information. This relatively late development is the result of a combination of factors including the need to place gamma ray telescopes above the Earth's atmosphere, the requirement to develop rather complex instruments, and the relatively low intensity of gamma ray photons particularly in relation to the charged particle cosmic ray intensity. It is worth noting that even though the photon intensity is low, the energy emitted in the gamma ray range may be, and in several cases is, quite high because each photon carries a large energy, and the gamma ray frequency range is very broad.

The first certain detection of high energy celestial gamma rays was made with the gamma ray telescope flown on the third Orbiting Solar Observatory

(OSO-3) by Clark, Garmire, and Krqushaar [1968], who observed gamma rays with energies above 0.5×10^8 eV from the galactic disk, with a peak intensity toward the galactic center. The galactic center emission was confirmed, and its narrow width measured with a large balloon-borne gamma ray telescope flown in 1969 [Kniffen and Fichtel, 1970]. About this same time, there were several reports of a high energy pulsed flux from the Crab in phase with the radio pulsar [Browning, Ramsden, and Wright, 1971; Albats et al., 1972; Parlier et al., 1973; McBreen et al., 1973; Helmken and Hoffman, 1973; and Kinzer, Share, and Seeman, 1973], as well as the first hints of very high energy ($\sim 10^{12}$ eV) gamma rays [Grindlay, 1972; Fazio et al., 1972].

On November 15, 1972, a gamma ray telescope with approximately 12 times the sensitivity of the OSO-3 instrument and angular resolution of one to two degrees was launched on the second Small Astronomy Satellite (SAS-2). This instrument provided results which led to a much better picture of the high energy (E>35 MeV) gamma ray sky including fair detail on the galactic plane [e.g., Fichtel et al., 1975; Bignami et al., 1979] and energy spectral and isotropy measurements on the diffuse extragalactic high energy gamma radiation [Fichtel et al., 1977]. Another gamma ray instrument (E>50 MeV) with approximately equal sensitivity and angular resolution carried by the Cosmic Ray Satellite (COS-B), was launched on August 8, 1975, and provided information which further expanded our knowledge [e.g., Mayer-Hasselwander et al., 1980 and 1982], including the first detection of a quasar in gamma rays. These data were supplemented in the medium energy range by instruments carried on high altitude balloons [e.g., Agrinier et al., 1981; Graser and Schonfelder, 1982; and Bertsch and Kniffen, 1983].

These data showed the rich character of the galactic plane diffuse emission with its potential for the study of the forces of change in the galaxy, the study of the origin and expansion of the cosmic ray gas and the study of the galactic structure. When examined in detail the longitudinal and latitudinal distribution appear generally correlated with galactic structural features, including spiral arm segments. Two molecular clouds have already been seen. With the observations of discrete sources, some of which are associated with supernovae and pulsars and others apparently not correlated with radiation at other wavelengths, point-source gamma ray astronomy has also begun. The Crab nebula exhibits both continuum and double-pulsed radiation with both pulses

in phase with the radio pulsar, PSR 0531 + 21. The Vela pulsar, PSR 0833 − 45, on the other hand has two pulses in the gamma ray region, as opposed to one in the radio region, neither of them in phase with the radio pulse [Thompson et al., 1975 and 1977a; Bennett et al., 1977; and Kanbach et al., 1980]. One of the two strongest gamma ray sources (l = 195, b = +5) yet observed [Thompson et al., 1977b; Swanenburg et al., 1981] had no obvious counterpart at other wavelengths at the time of its discovery by SAS-2. The highly variable X-ray source Cygnum X-3 was seen by SAS-2 [Lamb et al., 1977], but not COS-B.

In the very high energy (> 10^{11} eV) region of the gamma ray spectrum, ground-based Cherenkov light reflector telescopes have evidence now of gamma ray emission not only from the Crab Pulsar PSR 0531 + 21, but also from the Vela Pulsar PSR 0833-45, Cygnus X-3, Centaurus-A, and possibly other sources. [For a summary, see, for example, Grindlay, 1982 or Weekes, 1983.] The implications with regard to the sources of these extremely energetic photons are obviously impressive.

In this paper, the current status of galactic extragalactic gamma ray astronomy will be summarized. Subjects of the lower end of the gamma ray spectrum which are covered elsewhere in this book, namely gamma ray spectroscopy and bursts, will not be emphasized. Finally, a short description of forthcoming gamma ray missions will be presented including the envisioned scientific significance of the data they should obtain.

2. GALACTIC GAMMA RADIATION

As noted in the introduction, the gamma ray sky is dominated by radiation from the galactic plane, which is generally assumed to be the sum of diffuse radiation and unresolved point sources. The diffuse radiation was well anticipated. As early as 1952, Hayakawa [1952] noted the effect of meson-producing nuclear interactions between cosmic rays and interstellar gas. In the same year Hutchinson [1952] discussed the production of bremsstrahlung radiation by cosmic rays. Even earlier, Feenberg and Primakoff [1948] examined the astrophysics significance of the Compton effect in regard to cosmic ray electrons. The study of the diffuse radiation and its implications for our

galaxy is complicated by the fact that the point source contribution for the most part appears diffuse to the high energy gamma ray satellite instruments that have flown thus far because the angular resolution of these instruments for individual photons has been only one to a few degrees, or poorer, depending on energy. It is difficult to estimate the point source contribution since so little is known about them with only a few having been clearly identified. Several factors, however, suggest that point sources may not be a major contributor [e.g., Cesarsky, 1980]. These include the uniformity of the energy spectrum of the observed galactic radiation over the plane and its distribution being about what is expected from cosmic ray interactions. Even the earliest SAS-2 galactic gamma ray results [Kniffen et al., 1973] showed the general correlation of the gamma radiation with galactic structure, and subsequent work has shown it in greater detail, even quantitatively as will be shown in this section. This diffuse gamma radiation of cosmic ray origin will now be discussed followed by a summary of the current status of the knowledge of point sources.

A. Cosmic Ray Interactions with Galactic Matter and Photons

Of the products formed in cosmic ray nucleon interactions with matter, the ones of most immediate interest to high energy gamma rays are the mesons, and among these are the most commonly produced ones, namely the π mesons. Many of the other mesons and hyperons also decay rapidly into π mesons. The charged π mesons decay into neutrinos and electrons adding to the cosmic ray electrons already present which also interact to produce gamma rays. The π° mesons, however, decay into two gamma rays which have equal energy in the rest frame. With a collection of π°'s of various energies, a gamma ray energy spectrum results with a maximum at approximately 68 MeV and a curve which is symmetric when plotted as a function of ln E_γ. The spectrum predicted for cosmic ray interactions with interstellar matter has nearly this shape being primarily slightly broader due to the minor components. For a further discussion of this interaction process and the ones about to be discussed, see Fichtel and Trombka [1981] and the references therein.

As the high energy cosmic ray electron primaries and secondaries interact with galactic matter, they produce gamma rays through bremsstrahlung. These gamma rays have a spectrum which reflects that of the electrons being approximately a power law.

Cosmic ray electrons also interact with starlight photons, for which both the optical and infrared ranges are important, and with the blackbody radiation to produce Compton gamma rays. The source functions of these interactions are very much smaller in the galactic plane in the vicinity of the solar system. The total contribution to the galactic gamma radiation, however, is not completely negligible because the cosmic ray and stellar photon scale heights above the galactic plane are much greater than those of the matter and, of course, the blackbody photon density is uniform. Hence, the integral intensity along a line of sight is closer to that of the bremsstrahlung than the source functions would imply.

As an example of the spectrum which results from these three interaction processes, the calculated energy spectrum of the galactic gamma radiation for a region near the galactic center taken from the work of Fichtel and Kniffen [1984] is shown in Figure 1 and compared to data. The spectral shape elsewhere in the galactic plane is nearly the same with only minor variations resulting from small differences in the relative number of secondary electrons and the relative contribution of the Compton electrons. The agreement between theory and observation on the shape of the spectrum adds to the factors mentioned earlier for there at least being good reasons for believing that the majority of the measured apparently diffuse radiation is probably the result of cosmic ray interactions.

1. Large Scale Diffuse Galactic Features

High energy gamma rays provide us the best known opportunity to determine the cosmic ray density distribution in the galaxy in general, in terms of spiral arms, and in association with molecular clouds. Thereby, they can ultimately contribute to our understanding of the dynamical effects in our galaxy in an important way. However, it is first necessary to have a good understanding of the galactic matter distribution, which at present is not as well-defined as one would like. It is, therefore, worth reviewing this subject to locate wherein the primary difficulty lies, since, as will be seen, it is an important factor in the current uncertainty in the large scale cosmic ray density distribution.

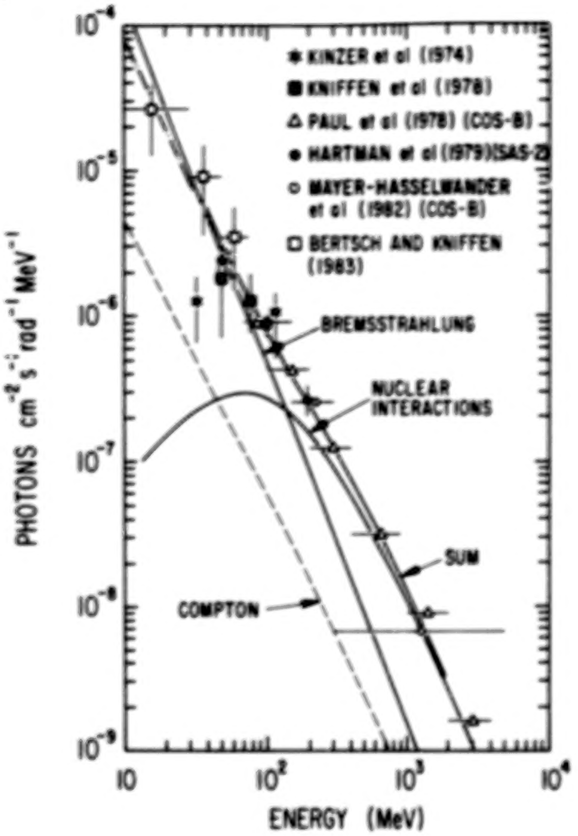


Figure 1. Energy spectrum of the galactic gamma radiation for a region near the galactic center. The calculated spectra are based on the work of Fichtel and Kniffen [1984]. The dot-dash curve includes an estimated correction for the increased energy loss by electrons in the inner galaxy. The 300-5000 MeV point of COS-B [Mayer-Hasselwander et al., 1982], which covers a large range in energy, is plotted at an energy where the differential energy spectrum of the equivalent power law spectrum is equal to the integral intensity divided by the energy interval width. The Compton component shown as a lightly dashed line is seen to be small and is somewhat uncertain. [This figure is from Fichtel and Kniffen, 1984.]

For the problem being considered here, the relevant concern is the galactic diffuse matter in the form of atoms, molecules, ions, and dust. The latter two are believed to be minor constituents and, hence, unimportant for gamma ray production through cosmic ray interactions. Hydrogen is the primary component of both the atomic and molecular matter. Helium and heavy nuclei add about 55% more to the gamma ray production. It is assumed these latter nuclei have a distribution in the galaxy similar to hydrogen, although little is known about them. Both atomic and molecular hydrogen are known to be confined to a narrow disk with the molecular hydrogen distribution generally having a smaller scale height [e.g., Gordon and Burton, 1976; Solomon and Sanders, 1980].

The neutral atomic hydrogen density distribution as revealed by the 21 cm emission is reasonably well-known; however, even it remains somewhat uncertain in the inner galactic regions because of uncertainty in the absorption correction. Recent work [e.g., Dickey et al., 1982; Thaddeus, 1982] suggests that the absorption had previously been somewhat underestimated and that the density in the region of 3-5 kpc from the galactic center is probably greater than previously estimated. The density distribution of molecular hydrogen is not as well-known because it is measured less directly. At present, the best estimate is obtained through the observation of the 2.6 mm spectral line of ^{12}CO, from which the distribution of cold interstellar matter is inferred. The nature of the interpretation of these measurements makes the derived molecular hydrogen density distribution less certain than that of the atomic hydrogen. In the long term, careful comparisons of gamma ray data with the atomic and molecular distributions should at least aid in the molecular hydrogen density normalization. The average galactic radial distributions of molecular and atomic hydrogen show clearly that the molecular hydrogen to atomic hydrogen ratio is larger in the inner galaxy than it is in the outer galaxy even if the absolute intensity of molecular hydrogen is still fairly uncertain. It is interesting to note that the great majority of the molecular hydrogen is in clouds. In most current analyses, the normalization of the molecular hydrogen density is treated as an adjustable parameter as long as it falls within the rather broad constraints set by other considerations.

If it were true that the cosmic ray density were constant throughout the galaxy, it would only be necessary to know the column density of the hydrogen in

order to calculate the diffuse galactic gamma ray emission [see, for example, Fichtel and Kniffen, 1974]. However, if the cosmic ray density is variable, and this question will be addressed below, the product of the cosmic ray density and the matter density must be integrated over the line of sight in the galaxy, and hence, the matter distribution in the galaxy must be deduced. Although the translation of the observations into a galactic spatial distribution is difficult, on a broad scale the density profile is reasonably well accepted. Even though there is not complete agreement on details of arm structure, a general spiral pattern does appear to emerge. In addition to the 21 cm data, the distributions of continuum radiation [Landecker and Wielebinski, 1970; Price, 1974], gamma radiation [Bignami et al., 1975], H II regions [Georgelin and Georgelin, 1976], supernova remnants [Clark and Caswell, 1976], pulsars [Seiradakis, 1976], and infrared emission [Hayakawa et al., 1976] are all consistent with the existence of spiral structure in the galaxy. Until recently, it had not been clear whether molecular clouds were associated with spiral structure. However, now on the basis of a high sample survey and observations in both the first and second quadrants of the galactic plane, Cohen et al. [1980] have reported the existence of the molecular counterparts of the five classical 21 cm spiral arms segments in these quadrants, namely the Perseus arm, the Local arm, the Sagittarius arm, the Scutum arm, and the 4 kpc arm.

With regard to the cosmic ray density distribution in the plane and perpendicular to it, there is some relevant experimental evidence and several theoretical considerations. The radio continuum measurements of Cane [1977] used together with the assumption that the galactic magnetic fields energy density and the cosmic ray energy density have the same scale height give a scale height for the cosmic rays of about 0.6 kpc relative to the plane of the galaxy. The galactic magnetic fields and the cosmic rays tied to these fields can only be constrained to the galactic disk by the gravitational attraction of the matter [Biermann and Davis, 1960; Parker, 1966, 1969, and 1977]. The local energy density of the cosmic rays (~ 1 eV/cm^3) is about the same as the estimated energy density of the magnetic field and that of the kinetic motion of matter. Together, the total expansive pressure of these three effects is estimated to be approximately equal to the maximum that the gravitational attraction can hold in equilibrium. Further, the cosmic ray age determination suggests that this situation is the result of plentiful sources and leakage, not just chance accumulation to the maximum over time. Hence, excluding the possibility that

the local conditions are anomalous, the most natural assumption is that the cosmic ray pressure is as great as it can be throughout the galaxy except possibly in the outer galaxy where sources or regions of further acceleration may be rare. The assumption that the cosmic ray density not only varies throughout the galaxy, but specifically on the scale of the arms, is based not only on the natural scale of the arms, but also on the scale height of cosmic ray electrons perpendicular to the plane, ~ 600 pc, and the theoretically suggested mean diffusion length in the plane (a few to several tenths of a kiloparsec). Furthermore, support for this assumption is obtained from the recent work showing that the cosmic ray electron intensity within the spiral arms is about a factor of 2 higher than between the arms [Webber, 1983].

Before considering the general case of the whole galaxy, two simpler examples for diffuse gamma ray production by cosmic rays in the galaxy will be noted. For galactic latitudes where the local contribution may be expected to dominate, $|b|$ greater than approximately 10°, the cosmic ray density as a function of position in the galactic plane presumably does not vary much. For this case, since the scale height of the cosmic rays is expected to be large compared to that of matter, a good approximation for the cosmic ray-matter interaction contribution to the gamma ray diffuse radiation is presumably obtained by using a constant cosmic ray density, which allows the direct use of atomic and molecular hydrogen column densities. If the point source contribution is small and if account is taken of the Compton contribution, it should be possible to obtain a good agreement using the matter column densities directly as shown by Strong et al. [1982] and Lebrun et al. [1982]. As an example, see Figure 2. It should also be possible to use this simplified approach successfully at intermediate longitudes, (~ 40° to ~ 120° and ~ 240° to ~ 320°), where regions which are at galactic radii similar to the Earth are predominantly being viewed as shown, for example, by Arnaud et al. [1982] and Lebrun et al. [1983]. See Figure 3.

The more general case wherein the cosmic ray density is allowed to be variable and specifically proportional to matter on the scale of galactic arms is treated by Fichtel and Kniffen [1984]. They show that within uncertainties the present gamma ray results are in agreement with a coupling of the cosmic ray density in the plane with the broad spiral arm scale galactic features as predicted by theory. Their results are shown in Figures 4 and 5. Notice that the edges

ORIGINAL PAGE IS
OF POOR QUALITY

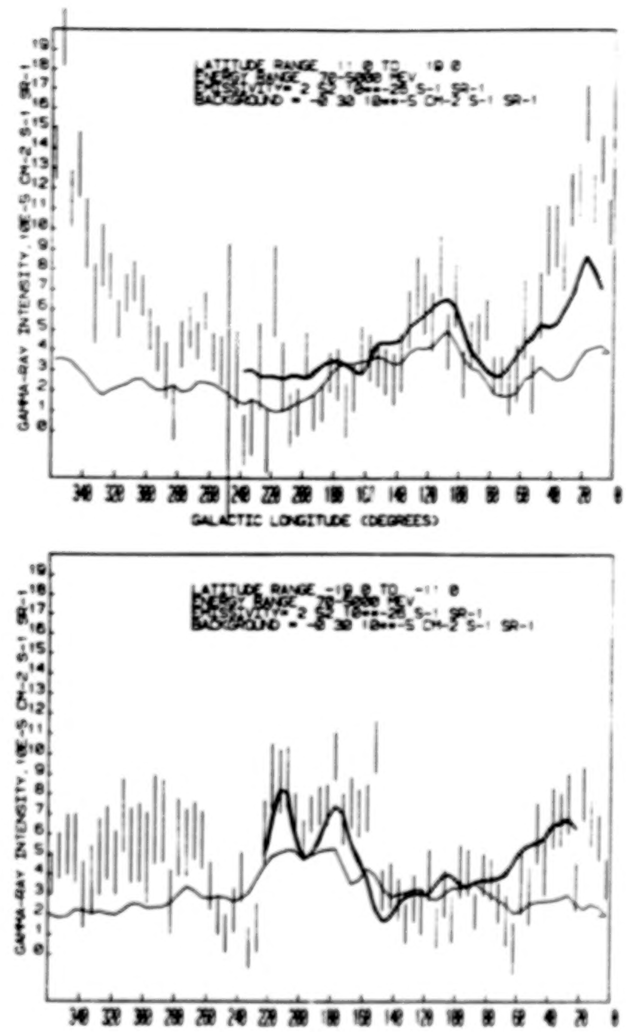


Figure 2. Comparison of measured and predicted gamma ray intensities. Bars: average measured intensity by COS-B with statistical errors based on counts in the bin. Thick line: predicted intensity based on estimated total column density (using galaxy counts). Thin line: predicted intensity for atomic hydrogen alone. [This figure is from Strong et al., 1982.]

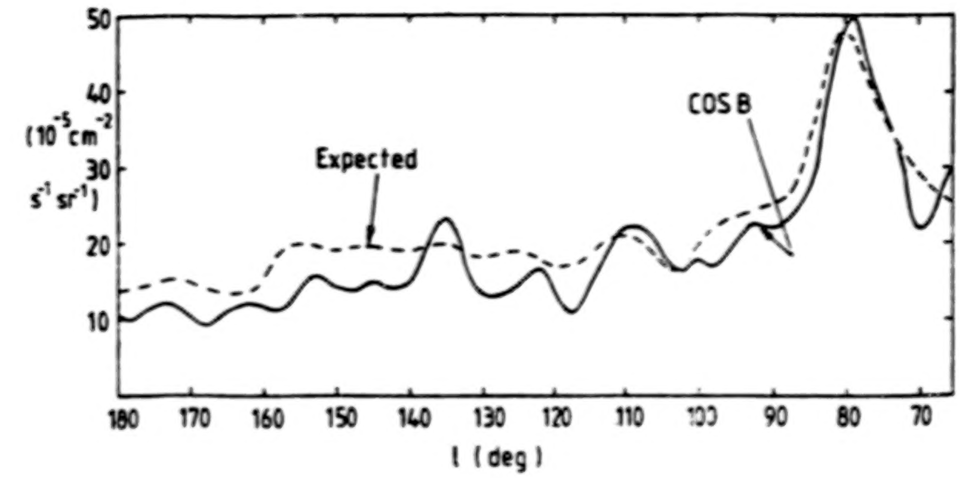


Figure 3. A comparison of the gamma ray intensity (E_γ>100 MeV) on a longitude along b = 0° between that observed by COS-B [Mayer-Hasselwander et al. 1982] and that calculated by Arnaud et al. [1982] under the assumption of a constant cosmic ray density everywhere at the local galactic value. [This figure is from Arnaud et al., 1982.]

of the Sagittarius and Crux arms at about 55° and 310° respectively mark the beginning of the higher intensity associated with the central region of that galaxy, and that further steps near 35° and 330° mark the edges of the Scutum and Norma arms. There appear to be increases at 75° and 285° associated with the local arm and the Carina arm respectively. The expected increase at 265° for the local arm is masked by the large increase due to the Vela pulsar. The latitude distributions are also generally reasonable. Fichtel and Kniffen [1984] obtain a normalization for molecular hydrogen of 1.3×10^{20} mol cm^{-2} K^{-1} Km^{-1} s^{-1}, which is consistent with the rather broad range allowed by independent analyses of radio and other data by Blitz and Shu [1980] and Dame and Thaddeus [1985].

Other analyses which support at least a general galactic radial gradient of the cosmic ray density include those of Dodds, Strong, and Wolfendale [1975], Kniffen, Fichtel, and Thompson [1977], Issa et al. [1981], Hermsen and Bloemen [1982], Bloemen et al. [1984], and Harding and Stecker [1985]. The

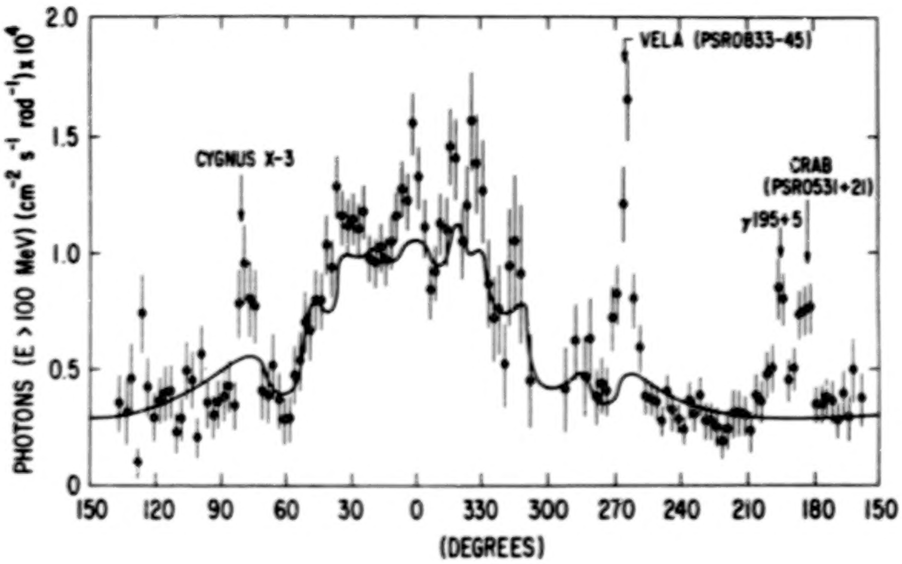


Figure 4. The high energy (E >100 MeV) gamma ray intensity as a function of longitude for $-10°<b<10°$ from the SAS-2 data [Hartman et al., 1979]. [This figure is from Fichtel and Kniffen, 1984.]

latter specifically deduce enhancements near the major inner galactic arm features. For example, an assumption of a constant cosmic ray density predicts too large a diffuse gamma ray intensity in the anticenter direction by a ratio of about 4:3 to 3:2, as noted several times before, e.g., Houston and Wolfendale [1982]. This result is expected, since as the matter density decreases in the outer galaxy, the cosmic ray density must also, since there is not then sufficient gravitational attraction to hold the local cosmic ray density.

It should be noted that Lebrun et al. [1983] showed that the gamma radiation above 300 MeV for ($12° < l < 97°$) and ($-5° < b < 10°$) may be fit by a linear combination of column densities of HI and CO gas tracers and an isotropic background. However, as Pollack et al. [1985] note, to interpret this result as implying a constant cosmic ray density is an oversimplification. Fichtel and Kniffen [1984] point out several important factors. First, since the molecular hydrogen density is concentrated towards the galactic center, a large normalization value for molecular hydrogen is mathematically similar

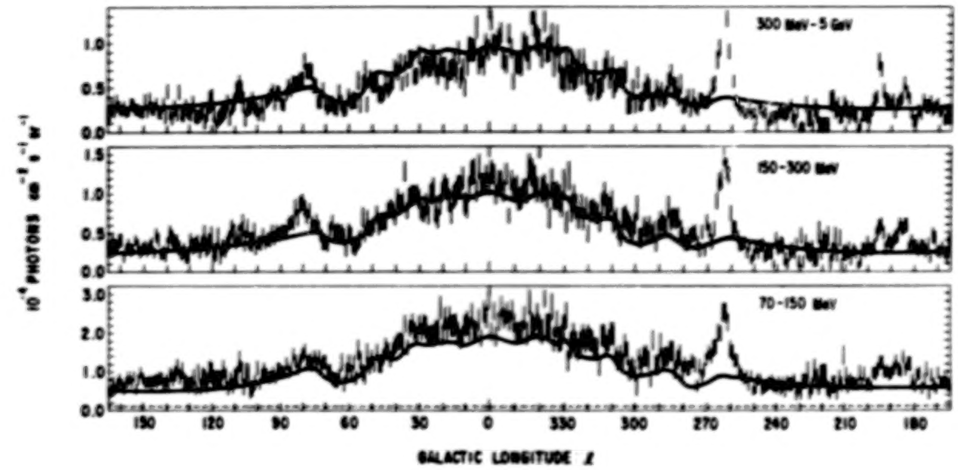


Figure 5. Gamma ray intensity as a function of longitude averaged over the latitude range $-10°<b<10°$ *from 70 MeV to 150 MeV, 150 MeV to 300 MeV, and 300 MeV to 5000 MeV from the COS-B data [Mayer-Hasselwander et al., 1982] compared to the model discussed here shown by the solid line. The dashed line in the 70-150 MeV graph represents the new "0.0" line based on the revised background intensity for the COS-B data in this energy interval. [This figure is from Fichtel and Kniffen, 1984.]*

to assuming a positive cosmic ray gradient towards the center in terms of the gamma rays produced. Lebrun et al. [1983] used 3.1×10^{20} molecules cm^{-2} K^{-1} Km^{-1} s. Second, the variation in the observed gamma ray intensity between 50° and 10° in l is less for 300 MeV $<E<$5000 MeV than in the other two COS-B energy ranges, for reasons yet to be determined. Third and least, the Compton contribution, as a percentage of the total, probably decreases toward the galactic center.

In summary, most evidence at present seems to support, or be consistent with, the theoretically supported concept of the correlation of the cosmic ray density with matter density on the broad scale of galactic arms. The anticenter region seems to show a decrease in cosmic ray intensity, while the inner galaxy is consistent with a higher cosmic ray density in the region of the galactic arms.

In addition to the galactic arms, the local galactic feature Gould's Belt has also been seen in gamma rays, having first been observed by Small Astronomy Satellite SAS-2 [Fichtel et al., 1975]. Also, there is the interesting feature of an apparent paucity of gamma rays coming from the galactic center itself (i.e., within a radius of 300 or 400pc of the center). Either the cosmic ray density must be very anomalously low relative to the matter density or the gas-to-dust ratio is exceptionally low. Based largely on circumstantial evidence, Blitz et al. [1985] favor a low H_2/CO abundance in the center region.

2. Molecular Clouds

Galactic molecular clouds contain much of the diffuse galactic matter and are generally believed to be the location for the formation of stars. There are, however, many unanswered questions about the nature of these regions and their formative processes. Gamma ray astronomy has the potential of making an important contribution through the determination of the cosmic ray distribution in these clouds. The limited sensitivity and angular resolution of the SAS-2 and COS-B instruments has restricted the gamma ray information on these objects; however, two positive identifications were made with the COS-B telescope, namely ρ Oph and the Orion cloud complex.

The Orion cloud complex has been observed as an extended gamma ray source [Caraveo et al., 1980] with the intensity distribution similar within uncertainties to the estimated matter distribution. See Figure 6. The centroids of the two excesses located approximately at α = 5h 4m, $\delta = 0°0'$; ($l = 205.4°$, $b = -14.4°$) and at α = 5h 30m, $\delta = -6°30'$ ($l = 210.0°$, $b = -20.6°$) are far enough from the galactic plane that source confusion is not likely. Detailed contours of the gamma ray emission, the CO line emission, the HI column densities, and the total gas column density are given separately by Bloemen et al. [1984], and then differences are taken to show a generally good correlation between the gamma radiation and the mass distribution.

The ρ Oph gamma ray excess [Swanenburg et al., 1981] is also clearly established, and its association with the cloud seems quite probable. There is a controversy over whether or not the gamma ray intensity is more than

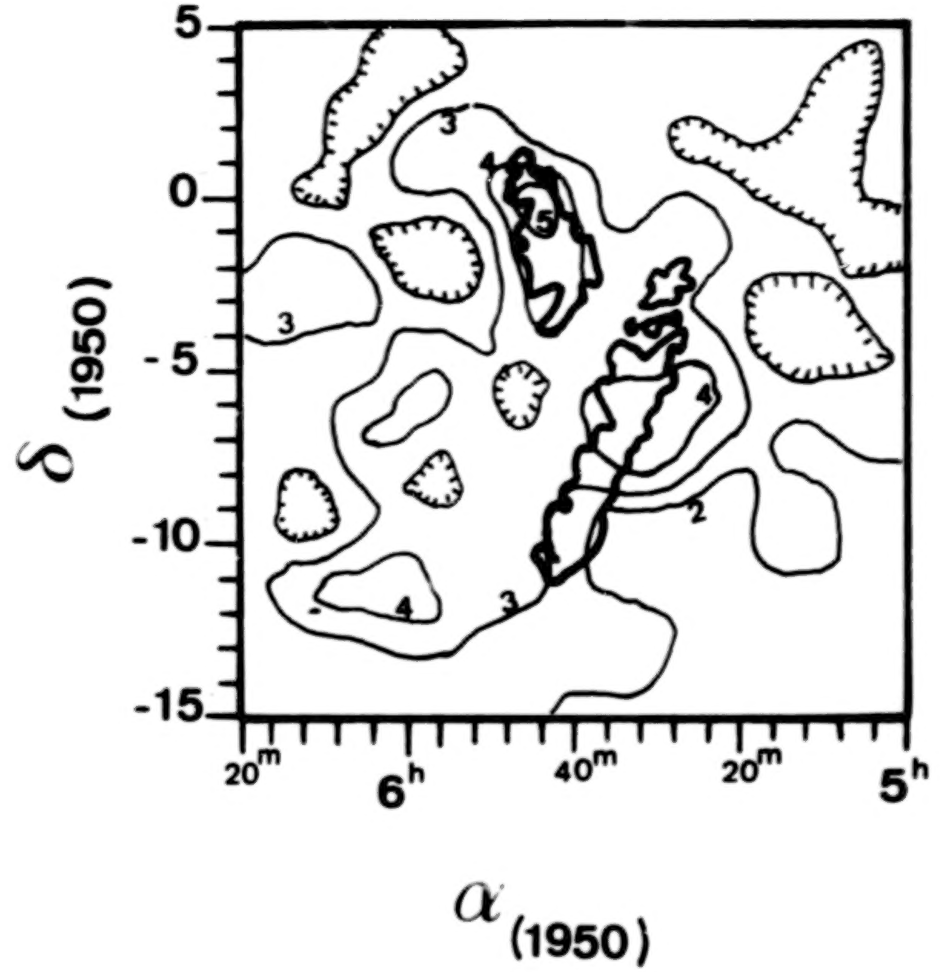


Figure 6. Gamma ray contour map of the Orion region. A smoothing has been applied to the data following Mayer-Hasselwander et al. [1980]. The contour unit is 2×10^{-3} "on axis counts" s^{-1} sr^{-1}. The thick line follows the boundaries of the L1641 and L1630 dark clouds. When considering the angular resolution of the COS-B satellite, the coincidence between the gamma ray excess and the cloud complex is further stressed. [This figure is from Caraveo et al., 1980.]

would be expected on the basis of cosmic rays at the local intensity level interacting with the matter in the cloud [For a summary, see Lebrun, 1984]. Although improved gamma ray data both in terms of angular resolution and statistics are desired, the greater uncertainty at present appears to lie in the estimate of the matter.

As noted, with future gamma ray observations of many clouds with greater sensitivity and better angular resolution, it is hoped that a better understanding can be achieved regarding the role of cosmic rays in internal cloud processes.

B. Compact Objects

The interest in the study of compact galactic objects in the gamma ray region, particularly neutron stars and black holes, is clearly quite high since these energetic photons are predicted to be able to provide answers on the fundamental nature of these objects. However, the angular resolution of the high energy gamma ray telescopes flown thus far, typically ½° to 1°, has generally made association with specific objects impossible except for cases where time variation correlation could be used. Many of the approximately two dozen localized excesses listed in the COS-B catalog [Swanenburg et al., 1981] are likely to be associated with molecular clouds, but some are probably yet to be resolved compact objects. [For a further discussion of this subject, see Pollack et al., 1985.]

1. Neutron Stars and Pulsars

Almost immediately after their discovery, pulsars were proposed to be associated with neutron stars [Gold, 1968, 1969], and this relationship is now generally accepted. The large release of energy, the very fast period, and the remarkably small variation of the period seem to dictate that the pulsed radiation must be from a massive object of small size. The very short length of the individual pulses indicates that the size of the emitting region is associated with something substantially smaller than normal stellar dimensions. On the other hand, the periods in general are constant to one part in 10^8 or greater indicating a massive object rather than a plasma phenomenon. If the period

of the pulse is associated with a rotating body, then the object must be a neutron star rather than a normal stellar object because the surface cannot move faster than the speed of light. Further, the period is probably too short to be associated with an oscillating phenomenon.

No consensus yet exists with regard to the exact model for the pulsed radiation. Many theoretical models involve either synchrotron or curvature radiation in the high magnetic fields, often in association with the polar regions to give the beaming effect. There is still, however, a great variety of opinions regarding the details of the model, including the specific manner and location of the relativistic particle acceleration.

The highest intensity pulsar as observed at the Earth in the gamma ray region is the one in Vela, PSR 0833-45, for which $(1.2 \pm 0.2) \times 10^{-5}$ photons E> 100 MeV) cm^{-2} s^{-2} are seen [Thompson et al., 1975 and 1977b; Bennett et al., 1977; and Kanbach et al., 1980]. This pulsar as first seen by SAS-2 is shown in Figure 7. The two most striking features are the two gamma ray pulses as opposed to one in the radio region, and the fact that neither gamma ray pulse is in phase with the radio pulse. These features were confirmed by the data obtained later from the COS-B satellite gamma ray telescope [Buccheri et al., 1978]. They determined that the first gamma ray pulse followed the radio pulse [e.g., Komesaroff, Hamilton, and Ables, 1972] by 11.2 ± 0.4 ms. The period is 89 ms, and the time between the pulses is 38 ms. If this result were not enough to complicate attempts to find a satisfactory theoretical model, following the detection of PSR 0833-45 in gamma rays two peaks were found in the optical region by Wallace et al. [1977], neither one of which was in phase with either the gamma ray or the radio peaks as seen in Figure 8 wherein the COS-B gamma ray data are shown. In spite of many attempts to obtain a certain detection of pulsation in the X-ray wavelength range, none has yet been made.

The observational picture for the second strongest gamma ray pulsar PSR 0531 + 21 in the Crab nebula is much simpler to describe. This pulsar, which is faster than PSR 0833-45 and was the first gamma ray pulsar reported [Browning, Ramsden, and Wright, 1971], even though it is not the strongest, is seen with the double pulsed structure in the radio, optical, X-ray, and gamma ray

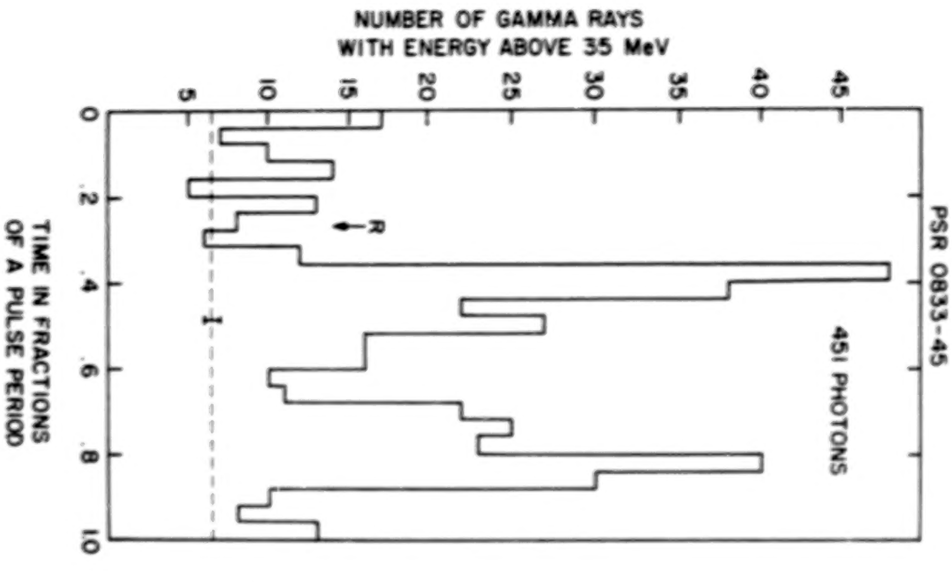


Figure 7. Distribution of gamma ray arrival times in fractions of a radio pulse period for gamma rays above 35 MeV from the direction of PSR 0833-45, as seen in the original data of the SAS-2 satellite. Arrow R marks the position of the radio pulse. The dashed line shows the gamma ray level expected from galactic and diffused radiation if no localized source were present [Thompson et al., 1977a]. Later data from PSR 0833-45 obtained with COS-B are shown in Figure 6. [This figure is from Thompson et al., 1977a.]

regions, and the pulses in each wavelength are in phase as shown in Figure 8. PSR 0531 + 21 has even been detected above 10^{11} eV by Helmken, Grindlay, and Weekes [1975] [see also, Grindlay, 1982; Cawley et al., 1985a; Kinov et al., 1985; and Tümer et al., 1985] using the ground-based 10 m reflector on Mt. Hopkins and others [e.g., Chadwick et al., 1985b; Cawley et al., 1985b]. The common in-phase double peak feature suggests the same mechanism for the radiation at all wavelengths for PSR 0531 + 21, which is perhaps to be expected for a pulsar which is only 10^3 years old. The older Vela pulsar, PSR 0833-45, apparently then has a dominant high energy component. In fact, whereas the pulsed luminosity ratio, L(PSR 0531 + 21)/L(PSR 0833-45), is about 5 above 100 MeV, it is almost 10^4 in the optical range. The

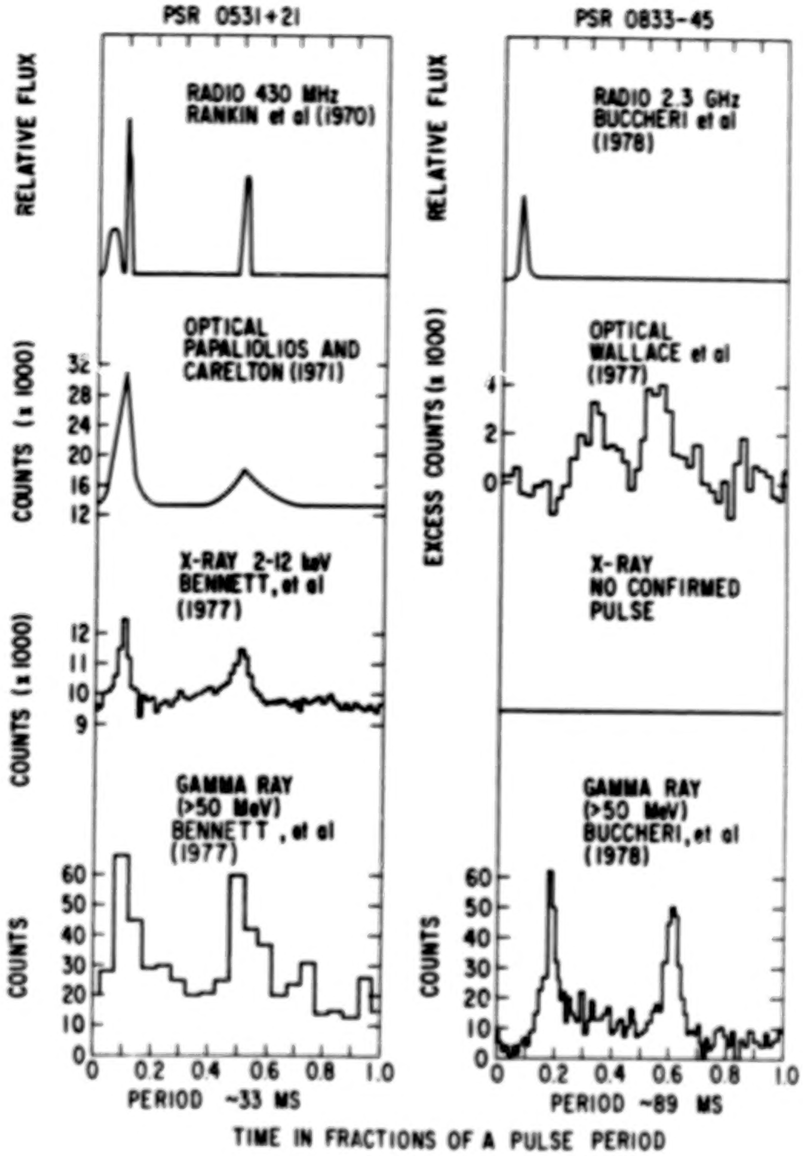


Figure 8. Comparison of the pulse structure and phase at radio, optical, X-ray, and gamma ray energies for PSR 0531 + 21 and PSR 0833 − 45, taken from Fichtel et al. [1980]. References to the work are given in the figure.

Crab nebula also has strong constant emission. The ratio of the pulsed to unpulsed emission appears to increase monotonically with energy until in the high energy gamma ray region the pulsed emission dominates.

The presence of 10^{12} eV gamma rays has been reported by Chadwick et al. [1985c] from the pulsar PSR 1953 + 29 with the characteristic 6.1 ms periodicity observed in the radio region. The 1.24s binary pulsar Her X-1 has also been seen at these very high energies [Dowthwaite et al., 1984; Cawley et al., 1985c].

Other radio pulsars have been reported to be possible gamma ray emitters both at 10^8 and 10^{12} eV, but confirmation is probably required because the statistical level of the observations is low. Most likely several of the radio pulsars will be revealed as gamma ray emitters. Semi-empirical analyses based on intensity as a function of age [Ogelman et al., 1976; Fichtel and Trombka, 1981] suggest that, with a factor of ten increase in sensitivity, many more pulsars will be seen in the gamma range.

2. Black Holes

Gamma rays are expected from black holes, although they are yet to be observed from a likely candidate. Maraschi and Treves [1977] have noted that, if the accretion flow onto the black hole is turbulent and dissipation maintains approximate equipartition among the different forms of energy, electrons can be accelerated by the induced electric fields. The resulting synchrotron energy spectrum is quite flat to about 20 MeV, above which it falls steeply. There would also be a Compton contribution. Under the right conditions, observable gamma ray fluxes would be generated. Collins [1979] has pointed out that matter falling onto a rotating black hole will be heated sufficiently that proton-proton collisions will produce mesons, including neutral pions which decay into two gamma rays. For massive ($> 10^3$ $M_{\odot}$) black holes, such as might exist in the galactic center, the resulting gamma ray luminosity may exceed 10^{36} ergs s^{-1}, which would give rise to over 3×10^{-6} gamma rays cm^{-2} s^{-1} at the Earth for a source at the galactic center. The energy spectrum would have a peak near 20 MeV. Emission from black holes through the Penrose process is another mechanism, but it is most appropriately discussed in relation to very large black holes which may exist at the centers of active galaxies.

One of the most intriguing predictions of gamma ray emission from black holes is that of bursts with a very unique signature from mini black holes left from the Big Bang [e.g., Page and Hawking, 1976; Hawking, 1977]. These mini holes, whose masses are only a very small fraction of that of the Sun, cannot be created in the universe as it exists today because the necessary compressional forces do not exist. As one of these primordial black holes continue to lose mass, its temperature rises, and it begins to emit particles of higher rest mass, until finally it ejects all its remaining rest mass in a very short time. The heavy hadrons emitted in this final release would decay very rapidly, giving about 10 to 30 percent of their energy ($\sim 10^{34}$ ergs) into a short burst (about 10^{-7}s) of hard gamma rays between 10^2 and 10^3 MeV, peaking around 250 MeV.

3. Other High Energy Gamma Ray Sources

Energetic ($E > 35$ MeV) gamma rays from Cygnus X-3 were observed with the SAS-2 gamma ray telescope by Lamb et al. [1977], as shown in Figure 9. They were modulated at the 4.8 hr period observed in the X-ray and infrared regions, and within the statistical error are in phase with this emission. The flux above 100 MeV had an average value of $(4.4 \pm 1.1) \times 10^{-6}$ photons $cm^{-2}\,s^{-1}$. Earlier, Galper et al. [1975] reported an excess above 40 MeV from Cygnus X-3 also with a 4.8 hr period at a higher intensity closer to the large 1972 radio burst, but with a lower statistical weight (3.6σ). If the distance to Cygnus X-3 is 10 kpc, the flux reported by SAS-2 implies a luminosity of more than 10^{37} ergs s^{-1} if the radiation is isotropic and about 10^{36} ergs s^{-1} if the radiation is restricted to a cone of one steradian, as it might be in a pulsar. At that luminosity level, during the time of the SAS-2 observation, Cygnus X-3 was the most luminous gamma ray source. However, since it is quite distant, the reported flux as observed at the Earth is just above the threshold for detection by SAS-2 or COS-B. When COS-B searched for radiation in later years (the SAS-2 observed it in 1973), no gamma radiation was detected [Bennett et al., 1977]. This result is not particularly surprising since the intensity as seen at other wavelengths such as X-rays was much lower then [e.g., Priedhorsky and Terrell, 1985]. The radio intensity is known to be quite variable.

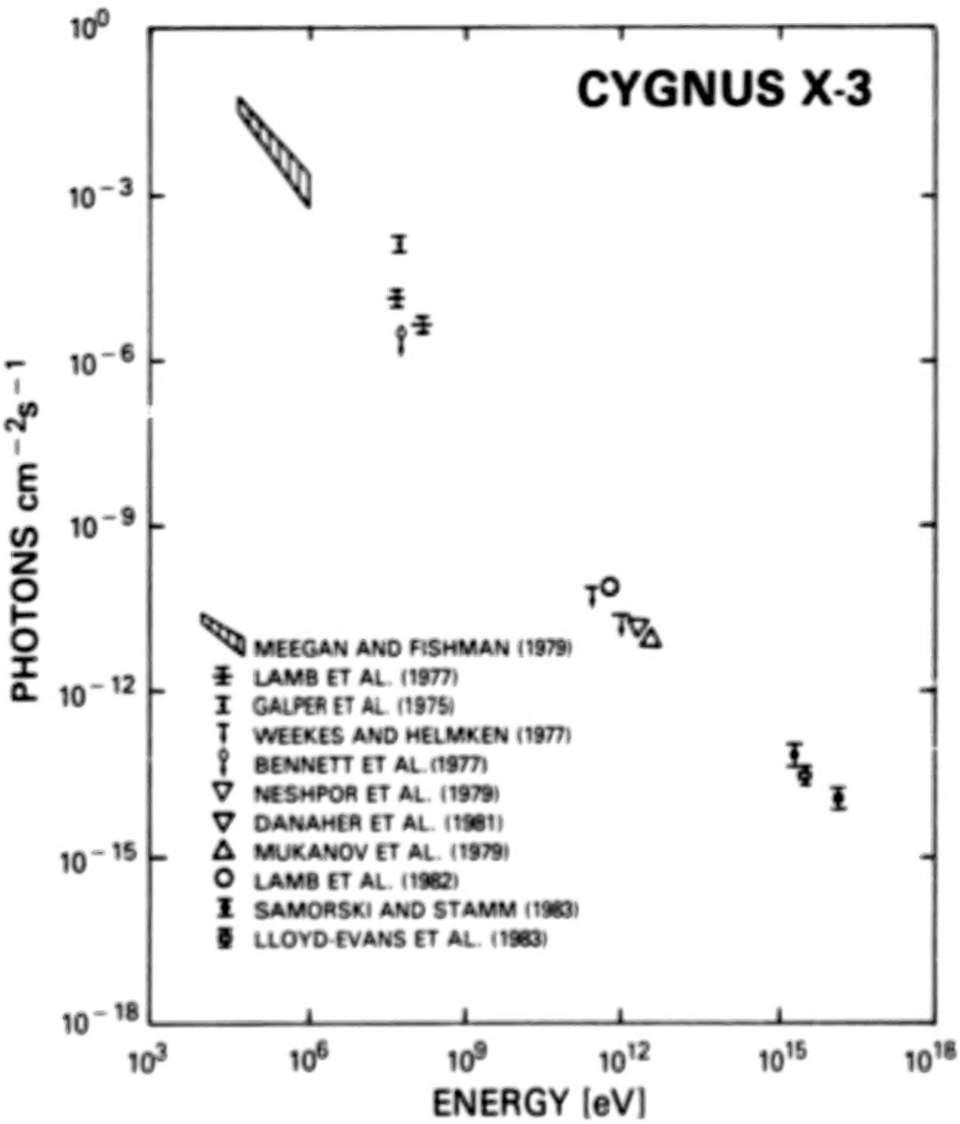


Figure 9. The photon energy spectrum observed for Cygnus X-3. References for the data points are shown in the figure. [This figure is a modification of one in Samorski and Stamm, 1983.]

There are also several reports of gamma ray emission from Cygnus X-3 at higher energies ($E > 10^{11}$ eV) [e.g., Vladimirsky, Stepanian, and Fomin, 1973; Stepanian et al., 1977; Danaher et al., 1981; Lamb et al., 1982; Dowthwaite et al., 1983; Cawley et al., 1985a; and Chadwick et al., 1985a], and even very high energies ($E > 10^{15}$ eV) [Samorski and Stamm, 1983; Lloyd-Evans et al., 1983; Dowthwaite et al., 1983; Kifune et al., 1985; Bhat et al., 1985; and Alexeenko et al., 1985], making it one of the more interesting astronomical objects ever observed. In the very high energy region, it is clear that once again Cygnus X-3 is not a steady emitter [e.g., Cawley et al., 1985d]. There are several null results as well as the positive results. These variations in intensity in the gamma ray region are consistent with the nonstatic theories of high energy particle acceleration in an accreting binary wherein there would be frequent breakdowns in the conditions.

There are several reasons for believing that the Cygnus X-3 gamma rays are produced by nuclei rather than electrons. Being massive, they are able to carry off energy acquired by acceleration in the high field region near the pulsar ($B > 10^5$ G) and deposit it in a low field region from which gamma rays are able to escape ($B < 10^3$ G). Electrons cannot do this [Hillas, 1984]. The production of gamma rays begins with the decay of the π° mesons formed in the collisions of accelerated nuclei with nuclei in the atmosphere of the companion star or in the gaseous material known to enshroud the Cygnus X-3 system. Thus, the energy spectrum of the accelerated nuclei must extend to at least 10^8 GeV per nucleon. It has been shown by Hillas [1984] that a monoenergetic beam of 10^8 GeV protons impinging on a stellar atmosphere will generate a cascade whose escaping gamma rays have a spectrum similar to that from Cygnus X-3. The acceleration could be driven by the rotational energy of the compact object [Eichler and Vestrand, 1984] or the accretion disk [Chanmugan and Brecher, 1985]. Alternately, the acceleration may be due to shocks [Kazanas and Ellison, 1986]. Fichtel and Linsley [1986] have noted that, in several of these models, there is more than enough energy being generated per unit time in the form of cosmic rays to sustain the observed intensity of $> 10^6$ GeV cosmic rays in the galaxy.

Supernovae, the most spectacular of stellar events, were among the first objects considered by gamma ray theorists. A continuum emission from the Crab

supernova remnant has been observed ranging upward to at least the low energy gamma ray region suggesting synchrotron radiation from relativistic electrons. Other supernovae have not been seen to have this type of radiation, but it may be only a question of sensitivity. A further important test of whether cosmic rays are accelerated by supernovae would be the observation of a continuum gamma radiation resulting from cosmic ray-matter interactions. A straightforward calculation assuming 10^{49} ergs in the form of relativistic particles and a density the same as that locally gives an intensity of about 10^{-7} photons ($E > 100$ MeV) cm^{-2} s^{-1} for a supernova 1 kpc away. This level would be detectable with future experiments, but not those flown thus far. For the Vela supernova remnant, which is closer, but in a possibly less dense region, a larger number might be predicted, making it an interesting candidate for future study. Also, supernovae in high density regions would be logical objects to search for high energy gamma rays [Montmerle, 1979].

As noted in the introduction, one of the two strongest high energy ($\sim 10^2$ MeV) gamma ray sources is the one at ($l = 195$, $b = +5$), which was first seen by SAS-2 [Lamb et al., 1977]. There is also a report of very high energy gamma rays from this source by Zyskin and Mukanov [1983 and 1985]. Radio, optical, and gamma ray searches have not revealed any object which stands out as being exceptional enough to have expected such strong gamma ray emission; however, there is now an X-ray source which seems to be a possible candidate [e.g., Bignami, Caraveo, and Lamb, 1983]. It is hoped that future gamma ray measurements may define the source location accurately enough to locate an object at lower wavelengths with more certainty.

3. EXTRAGALACTIC GAMMA RADIATION

A. Active Galaxies

Beyond our galaxy gamma ray emission has already been seen from four active galaxies, two Seyfert galaxies, one radio galaxy, and a quasar. For at least two of these, 3C 273 and NGC 4151, more energy is emitted in the gamma ray region ($E > 0.1$ MeV) than in the X-ray, optical, or radio regions. The photon intensities are relatively low, of course, because of the large energy per photon in the gamma ray region. No normal galaxy, other than our own, has been seen in gamma rays, but this result is not surprising on the basis

of the emission level of our own galaxy [See, for example, Fichtel and Trombka, 1981]. The gamma rays observed from the Seyfert galaxies NGC 4151 and MCG 8-11-11 are in the low energy gamma ray region [Auriemma et al., 1978; Coe et al., 1981; Gursky et al., 1971; Ives, Sanford, and Penston, 1976; Meegan and Haymes, 1979; Mushotsky, Holt, and Serlemitsos, 1978; Paciesas, Mushotsky, and Pelling, 1977; Perotti et al., 1979; Perotti et al., 1981; Schonfelder, Graml, and Penningsfeld, 1980; and White et al., 1980], and only upper limits to the high energy gamma radiation exist [Bignami et al., 1979]. The spectra are similar in that both show a very marked increase in the spectral slope in the energy region near 1 MeV. The several measurements in the gamma ray region for NGC 4151 were made at different times, and, assuming no significant errors in the data, clearly show a time variability. For five other Seyferts (and also several other emission-line galaxies) upper limits derived from the SAS-2 gamma ray data [Bignami et al., 1979] are substantially (more than an order of magnitude) below an extrapolation of the power law X-ray spectra [Mushotsky, Holt, and Serlemitsos, 1978], suggesting that a sharp spectral change in the low energy gamma ray region may be a general feature of these galaxies. Pollock et al. [1981] come to a similar conclusion in relation to the COS-B data.

Turning to quasars, 3C 273 is the brightest X-ray quasar and is the only quasar which has been clearly identified as a source of gamma rays [Swanenburg et al., 1978]. The differential energy spectrum of 3C 273 steepens sharply from the X-ray range to the gamma ray region, with the slope of the differential energy spectrum changing from 1.4 in the hard X-ray region to 2.7 in the high energy (E>50 MeV) gamma ray region as shown in Figure 10. The change in spectral shape between the hard X-ray and gamma ray region seen for 3C 273 is similar to that suggested for the Seyfert galaxies for which data exist. The COS-B instrument has observed 3C 273 in gamma rays in July 1976, June 1978, and June-July 1980, and no significant variation in the gamma ray fluxes among the observation was observed [Bignami et al., 1981b].

The closest known quasar is 2S 0241 + 622, but it is very close to the galactic plane ($b \simeq 2°$). The error box of the COS-B gamma ray source CG 135 + 1 [Hermsen et al., 1977] contains the position of 2S 0241 + 622, and the possible association has been pointed out by Apparao et al. [1978]. Because of

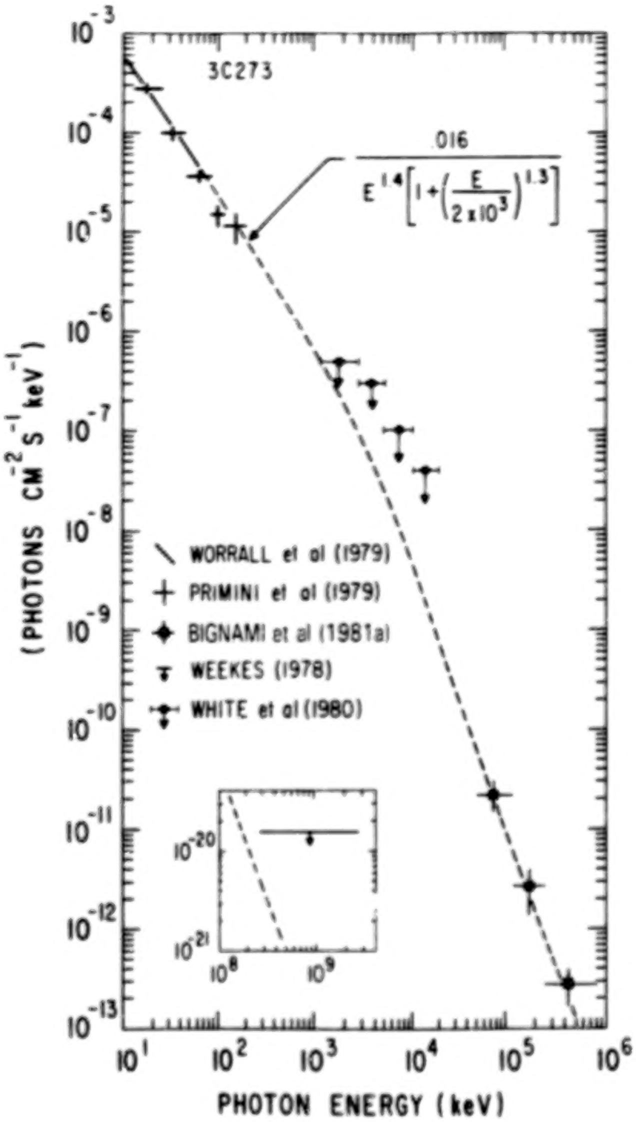


Figure 10. The differential high energy photon spectral results from several experiments as a function of energy for 3C 273. The dashed curve is given by the equation shown in the figure. [This figure is from Fichtel and Trombka, 1981.]

the large area of the gamma ray source within the galaxy, this identification must be considered tentative.

The change in spectral shape between the hard X-ray and gamma ray region seen for 3C 273 is similar to that suggested for the Seyfert galaxies for which data existed. This spectrum, although as yet poorly defined, is consistent with several of the massive black hole models, including the synchrotron, self Compton type models [e.g., Grindlay, 1975; Shapiro and Saltpeter, 1975; Mushotzky, 1976; and Maraschi and Treves, 1977], possibly including some degree of subsequent photon-photon interactions [Jelley, 1966] and the Penrose pair production and Compton scattering processes [Leiter and Kafatos, 1978; Kafatos and Leiter, 1979], involving infalling protons and electrons.

Centaurus-A (NGC 5128), generally believed to be the closest radio galaxy, is the only radio galaxy that has been seen in gamma rays. It has now been observed in all frequency bands from radio through low energy gamma rays and, although gamma ray emission is not seen in the 30 to 10^3 MeV region [Bignami et al., 1979; Pollock et al., 1981], a strong indication of very high energy ($E>3\times10^{11}$ eV) gamma ray emission has been found [Grindlay, 1975]. The observations of the radiation from CEN-A in the X-ray region through the very high energy gamma ray region again suggest a steepening of the spectral slope possibly similar to NGC 4151, MCG 8-11-11, and 3C 273.

B. Diffuse Extragalactic Radiation

A diffuse celestial radiation, which is isotropic at least on a coarse scale, has been measured from the soft X-ray region to at least 150 MeV. The first indication that diffuse celestial radiation extended from the X-ray region into at least the low energy gamma ray (1 MeV) portion of the spectrum was reported by Arnold et al. [1962]. At energies above 10 MeV, the first measurements related to an extragalactic diffuse radiation were those of Kraushaar and Clark [1962], whose upper limits from Explorer 11 provided an experimental refutation of the steady-state theory of cosmology. The first suggestion of a diffuse high energy flux came from the Orbiting Solar Observatory OSO-3 satellite experiment [Kraushaar et al., 1972], and it was data from the SAS-2 high energy gamma ray experiment that clearly established

a high energy extension of the diffuse radiation with a steep energy spectrum above 35 MeV [Fichtel et al., 1977]. A recent reanalysis of the SAS-2 data which included galaxy counts as a tracer of the interstellar matter has been performed by Thompson and Fichtel [1982] and has added support to the concept of the spectrum being quite steep (having a power low index of about 2.4) in the energy region above 35 MeV. See Figure 11.

Although the diffuse spectral measurements are reasonably self-consistent, the degree of spatial isotropy is not well known. The X-ray spectrum through about 100 keV is known [cf., Schwartz and Gursky, 1973] to be isotropic to within about 5 percent. At high gamma ray energies (35 to 100 MeV), the center-to-anticenter ratio for radiation with $20° < |b| < 40°$ was measured to be 1.10 ± 0.19 and the perpendicular to the galactic plane intensity to that in the $20° < |b| < 40°$ region was measured to be 0.87 ± 0.09 each of these results is consistent with isotropy to within errors [Fichtel, Simpson, and Thompson, 1978]. Although much more precise measures of the isotropy are clearly desired, no evidence for a major anisotropy exists. In particular, the high energy gamma ray results just quoted eliminate a spherical galactic halo origin for the radiation in view of the Sun's great distance from the galactic center. For the future, trying to establish the level of isotropy or deviations therefrom on both a coarse scale and a fine scale will be quite important.

A large number of theories predicting a diffuse gamma ray background have appeared in the literature over the years. With the measurements of the spectrum and intensity which now exist, most of these seem not to be likely candidates for the majority of the diffuse radiations [see, for example, Fichtel and Trombka, 1981]. Two possibilities seem to remain at present. One of these involves a baryon-symmetric universe, containing superclusters of galaxies of matter and others of antimatter. The annihilation of nucleons and antinucleons at the boundaries [Stecker, Morgan, and Bredekamp, 1971] produces the gamma rays. The predicted energy spectrum is reasonable and the required normalization is within the rather wide currently accepted range. This theoretical model predicts a fairly smooth distribution over the sky; however, a test of this theory (in addition to a precise measure of the energy spectrum) would be the detection of fairly small enhancements in the gamma radiation in the direction of boundaries between close superclusters of galaxies. The

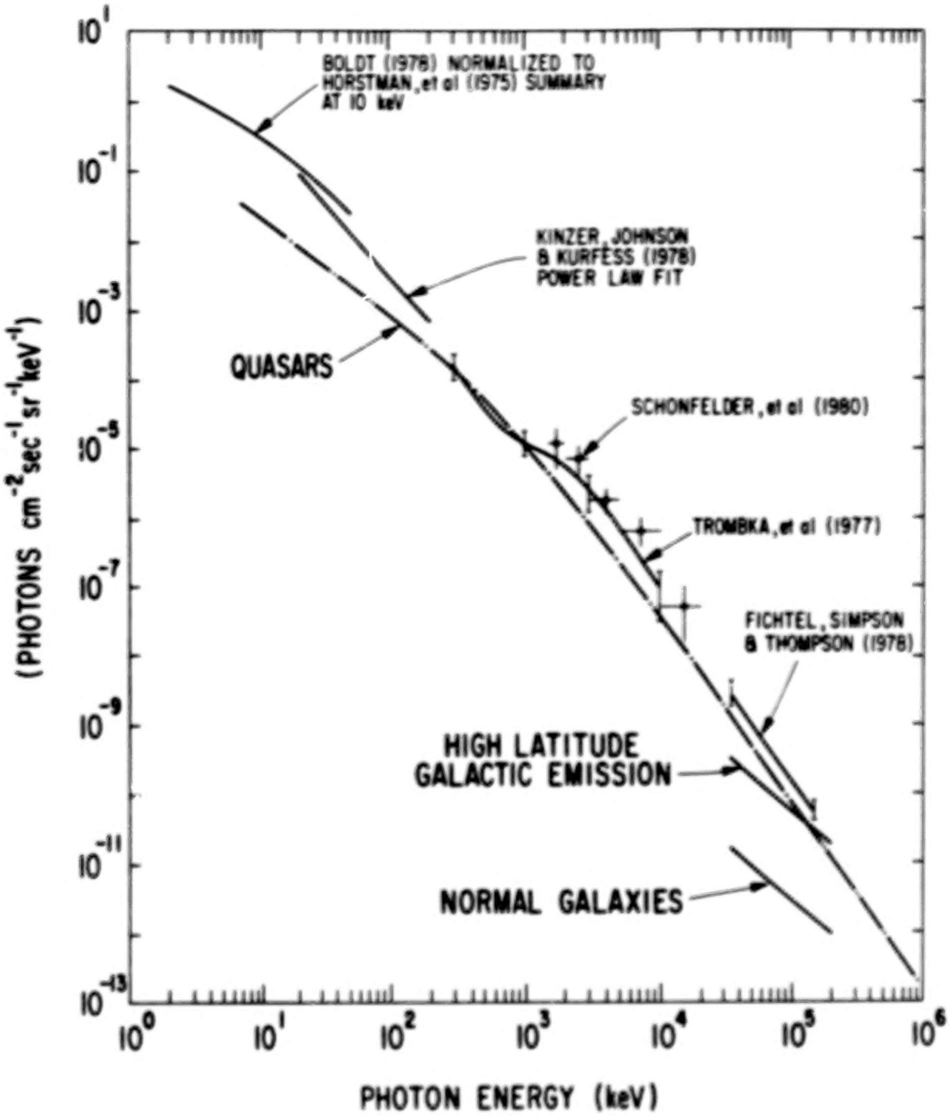


Figure 11. The diffuse gamma ray energy spectra estimated for quasars and normal galaxies under the assumptions described in the text, compared to some of the more recent experimental data. [This figure is from Fichtel and Trombka, 1981.]

diffuse gamma radiation associated with these particular boundaries would be at the higher energies.

The other possibility is the sum of the radiation from point sources, and particularly active galaxies, integrated over cosmological times [e.g., Strong and Worrall, 1976; Bignami, Lichti, and Paul, 1978; Schonfelder, 1978; and Grindlay, 1978]. Using the data on the few known objects, Bignami et al. [1979] and Fichtel and Trombka [1981] conclude that it is quite conceivable that Seyfert galaxies and quasars could account for the diffuse gamma radiation using conservative evolutionary models. The estimate of the quasar contribution by Fichtel and Trombka [1981], based on the existing, very limited knowledge is shown in Figure 11. Leiter and Boldt [1982] and Boldt and Leiter [1984] have proposed a model based on supermassive Schwarzschild black holes with accretion disks radiating near the Eddington luminosity limit. The authors believe that, if this theory is correct, there would be detectable variations in the diffuse radiation in small elements of the sky (10 deg^2) over several days in the ½ to 3 MeV region, giving a specific test for this theory.

4. FUTURE PROSPECTS

The satellite opportunities for high energy gamma ray astronomy in the near future are the GAMMA I planned to be launched in late 1986 and the Gamma Ray Observatory, currently scheduled for launch in 1988. In the most distant future, the NASA Space Station should provide opportunities to fly quite large gamma ray instruments which might be refurbished or reconfigured in space.

A. GAMMA I

The next gamma ray satellite expected to fly is GAMMA I. It is similar to SAS-2 and COS-B in the sense that its central element is a multilayer spark chamber system, triggered by a directional counter telescope, and surrounded on the upper end by an anticoincidence system. The upper spark chamber system is a twelve-level wide gap Vidicon system. The directionality of the electrons is determined by a time-of-flight system rather than a directional Cherenkov counter. The sensitive area is about 1600 cm^2 or about 2.7 times

that of SAS-2 or COS-B. The area solid angle factor is about the same, because the viewing angle is smaller. The gamma ray arrival direction measurements are expected to be an improvement over those of SAS-2 and COS-B.

B. The Gamma Ray Observatory (GRO)

Frank McDonald played a major role in bringing this observatory into being. Following his early work during the formative period, he was named the GRO Study Scientist for the initial formal phase of GRO's life. The Gamma Ray Observatory (GRO) is now an approved NASA mission with a launch tentatively planned for 1990, as mentioned above. An artist's concept of this spacecraft, which is to be launched by the shuttle, is shown in Figure 12. There are four instruments covering the energy range from 0.03 MeV to 3×10^4 MeV, with a major increase in sensitivity over previous satellite experiments. It is advantageous to combine the instruments into one mission not only because they place similar requirements on a spacecraft, but also because of the great scientific value of studying the entire gamma ray spectrum of any object at the same time to examine in detail the nature of time variations.

The four instruments to fly on GRO are:

1. Gamma Ray Observatory Scintillation Spectrometer Experiment (OSSE)

This experiment utilizes four large actively shielded and passively collimated sodium iodide (NaI) scintillation detectors, with a $5° \times 11°$ FWHM field of view. The large area detectors provide excellent sensitivity for both gamma ray line and continuum emissions. An offset pointing system modulates the celestial source contributions to allow background subtraction. It also permits observations of off-axis sources such as transient phenomena and solar flares without impacting the planned Observatory viewing program. The energy range is 0.1 to 10 MeV.

2. Imaging Compton Telescope (COMPTEL)

This instrument employs the signature of a two-step absorption of the gamma ray, i.e., a Compton collision in the first detector followed by total absorption in a second detector element. This method, in combination with effective charged particle shield detectors, results in a more efficient suppression

ORIGINAL PAGE IS
OF POOR QUALITY

Figure 12. Artist's concept of the Gamma Ray Observatory taken from a color print prepared by TRW.

of the inherent instrumental background. Spatial resolution in the two detectors together with the well-defined geometry of the Compton interaction permits the reconstruction of the sky image over a wide field of view (~ 1 steradian) with a resolution of a few degrees. In addition, the instrument has the capability of searching for polarization of the radiation. The instrument has good capabilities for the search for weak sources, weak galactic features, and for the search for spectral and spatial features in the extragalactic diffuse radiation. The energy range is 1 to 30 MeV.

3. Energetic Gamma Ray Experiment Telescope (EGRET)

The High Energy Gamma Ray Telescope is designed to cover the energy range from 20 MeV to 30×10^3 MeV. The instrument uses a multithin-plate spark chamber system to detect and record gamma rays converted by the electron-positron pair process. A total energy counter using NaI (Tl) is placed beneath the instrument to provide good energy resolution over a wide dynamic range. The instrument is capped with a plastic scintillator anticoincidence dome to prevent readout on events not associated with gamma rays. The combination of high energies and good spatial resolution in this instrument provides the best source positions of any GRO instrument.

4. Burst and Transient Source Experiment (BATSE)

The Burst and Transient Source Experiment for the GRO is designed to continuously monitor a large fraction of the sky for a wide range of types of transient gamma ray events. The monitor consists of eight wide field detector modules. Four have the same viewing path as the other telescopes on GRO and four are on the bottom side of the instrument module viewing the opposite hemisphere. This arrangement provides maximum continuous exposure to the unobstructed sky. The capability provides for 0.1 msec time resolution, a burst location accuracy of about a degree, and a sensitivity of 6×10^{-8} erg/cm^2 for a 10 sec burst.

The combined complement of instruments to be incorporated into the Gamma Ray Observatory is expected to have the capability to carry out the following:

i. A survey of gamma ray sources and diffuse emission with sensitivities around 10^{-5} photon cm^{-2} sec^{-1} and energy resolution around 10 percent at energies between 0.1 and 30 MeV.

ii. A survey of high energy gamma ray sources and diffuse emission with a point source sensitivity of 10^{-7} photons cm^{-2} sec^{-1}, angular resolution of about 0.1° for strong sources, and energy resolution around 15% at energies above 10^2 MeV.

iii. Detection and identification of nuclear gamma lines with an energy resolution of 4 percent and sensitivity of the order of 5×10^{-5} photon cm^{-2} sec^{-1}.

iv. Observations of gamma ray bursts, including studies of their spectral and temporal behavior.

The Gamma Ray Observatory will be a shuttle-launched, free-flyer satellite. The nominal circular orbit will be about 400 kilometers with an inclination of 28.5°. Celestial pointing to any point on the sky will be maintained to an accuracy of ± 0.5°. Knowledge of the pointing direction will be determined to an accuracy of 2 arc-minutes. Absolute time will be accurate to better than 0.1 milliseconds to allow precise comparisons of pulsars and other time varying sources with observations at other wavelengths from ground observations and other satellites. For further information on the instruments to be flown on GRO, see Kniffen et al. [1981].

C. Space Station

NASA is now considering a Space Station which would be a manned spacecraft permanently orbiting the Earth and capable of performing a variety of scientific missions. High energy gamma ray astronomy is certainly among the scientific disciplines which would be able to benefit very significantly from such an opportunity. With the gamma ray sky surveyed in some depth with the GRO, it would, for example, be possible to concentrate on the detailed features of discrete sources such as active galaxies to examine detailed characteristics including the temporal variability on time scales up to years and to study in depth limited regions such as clouds, galactic arms, and nearby galaxies.

D. Very High Energy Ground-Based Gamma Ray Telescopes

In the very high energy realm of gamma ray astronomy, there are plans for ground-based atmospheric Cherenkov experiments that would parallel the developments in sensitivity at gamma ray energies in the MeV-GeV region expected in the Space Station era. One of these [Weekes, 1985] consists of seven 10 to 15 m aperture optical reflectors in an array of spacing 7 m at a high mountain altitude (3.5 km). Each reflector would be equipped with a camera similar to that currently in use at the Whipple Observatory. The flux sensitivity will be a factor of ten better than that achievable with the current camera and will be competitive with the anticipated sensitivity of spaceborne calorimeters in the TeV energy range. The effective energy threshold could be as low as 10 GeV (10^{10} eV). A co-located particle detector array consisting of 61 scintillators of 1 m^2 area will give coverage in the even higher energy range from 10^{14} to 10^{17} eV, so that six magnitudes of the electromagnetic spectrum will be simultaneously monitored.

Should all these opportunities come into being, gamma ray astronomy should make a major contribution to the understanding of the universe.

REFERENCES

Agrinier, B. et al., 1981, *17th Internat. Cosmic Ray Conference Papers (Paris)*, 72.

Albats, P. et al., 1972, *Nature*, **240**, 221-24.

Alexeenko, V. V. et al., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **1**, 91.

Apparao, K. M. V. et al., 1978, *Nature*, **273**, 453.

Arnaud, K. et al., 1982, *Mon. Not. Roy. Astr. Soc.* **201**, 745.

Arnold, J. R. et al., 1962, *J. Geophys. Res.*, **67**, 4876.

Auriemma, G. et al., 1978, *Astrophys. J.*, **221**, L 7.

Bennett, K. et al., 1977, *Astron. Astrophys.*, **61**, 279.

Bertsch, D. L., and Kniffen, D. A., 1983, *Astrophys. J.*, **70**, 305.

Bhat, C. L. et al., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **1**, 83.

Biermann, L., and Davis, L., 1960, *Zs. F. Ap.*, **51**, 19.

Bignami, G. F. et al., 1975, *Space Sci. Instrum.*, **1**, 245.

Bignami, G. F., Lichti, G. G., and Paul, J. A., 1978, *Astron. & Astrophys.*, **68**, L15.

Bignami, G. F., Fichtel, C. E., Hartman, R. C., and Thompson, D. J., 1979, *Astrophys. J.*, **232**, 649.

Bignami, G. F. et al., 1981a, *Astron. Astrophys.*, **93**, 71.

Bignami, G. F. et al., 1981b, *17th Internat. Cosmic Ray Conference Papers (Paris)*, **1**, 182.

Bignami, G. F., Caraveo, P. A., and Lamb, R. C., 1983, *Astrophys,. J.*, **272**, L9.

Blitz, L., and Shu, F. H., 1980, *Astrophys. J.*, **238**, 148.

Blitz, L. et al., 1985, *Astron. Astrophys.*, **143**, 267.

Bloemen, J. B. G. M. et al., 1984, *Astron. Astrophys.*, **135**, 12.

Boldt, E. A., 1978, *The Diffuse Component of the Cosmic X-Radiation*, NASA TM #78106.

Boldt, E., and Leiter, D., 1984, *Astrophys. J.*, **276**, 427.

Browning, R., Ramsden, D., and Wright, P. J., 1971, *Nature*, **232**, 99-101.

Buccheri, R. et al., 1978, *Astron. Astrophys.*, **69**, 141.

Cane, H. V., "Non-Thermal Galactic Background Radiation." Ph.D. dissertation, University of Tasmania, Hobart, 1977.

Caraveo, P. A. et al. 1980, *Astron. Astrophys.*, **91**, L3.

Cawley, M. F. et al., 1985a, *Astrophys. J.*, in press.

Cawley, M. F. et al., 1985b, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **1**, 131.

Cawley, M. F. et al., 1985c, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **1**, 119.

Cawley, M. F. et al., 1985d, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **1**, 87.

Cesarsky, C. S., 1980, *Annals of the New York Academy of Sciences*, **336**, 223.

Chadwick, P. M. et al., 1985a, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **1**, 79.

Chadwick, P. M. et al., 1985b, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **1**, 155.

Chadwick, P. M. et al., 1985c, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **1**, 161.

Chanmugan, G., and Brecher K., 1985, *Nature*, **313**, 767.

Clark, G. W., Garmire, G. P., and Krqushaar, W. L., 1968, *Astrophys. J.*, **153**, L203.

Clark, O. H., and Caswell, L. L., 1976, *Mon. Not. Roy. Astr. Soc.*, **174**, 267.

Coe, M. J. et al., 1981, *Mon. Not. Roy. Astr. Soc.*, **195**, 24.

Cohen, R. S., Cong, H., Dame, T. M., and Thaddeus, P., 1980, *Astrophys. J.*, **239**, L53.

Collins, M. W., 1979, *Astron. Astrophys.*, **74**, 108.

Dame, T. M., and Thaddeus, P., 1985, *Astrophys. J.*, **297**, 751.

Danaher, S., Fegan, D. J., Porter, N. A., and Weekes, T. C., 1981, *Nature*, **289**, 568.

Dickey, J. M., Kulkarni, S. R., Van Gorkom, J. H., Benson, J. M., and Heiles, C. E., 1982, "Low Latitude Absorption Spectra," (National Radio Astronomy Observatory).

Dodds, D., Strong, A. W., and Wolfendale, A. W., 1975, *Mon. Not. Roy. Astr. Soc.*, **171**, 569.

Dowthwaite, J. C. et al., 1983, *Astron. Astrophys.*, **126**, 1.

Dowthwaite, J. C. et al., 1984, *Nature*, **309**, 691.

Eichler, D., and Vestrand, W. T., 1984, *Nature*, **307**, 613.

Fazio, G. G., Hemken, H. F., O'Mongain, E., and Weekes, T. C., 1972, *Astrophys. J.*, **175**, L117-L122.

Feenberg, E., and Primakoff, H., 1948, *Phys. Rev.*, **73**, 449.

Fichtel, C. E., and Kniffen, D. A., 1974, in *High Energy Particles and Quanta in Astrophysics*, ed. F. B. McDonald and C. E. Fichtel (Cambridge: The MIT Press).

Fichtel, C. E. et al., 1975, *Astrophys. J.*, **198**, 163.

Fichtel, C. E. et al., 1977, *Astrophys. J.*, **217**, L9.

Fichtel, C. E., Simpson, G. A., and Thompson, D. J., 1978, *Astrophys. J.*, **222**, 833.

Fichtel, C. E., Arnett, D., Grindlay, J., and Trombka, J., 1980, in *Astrophysics from Spacelab*, ed. P.L Bernacca and R. Ruffin (Dordrecht: D. Reidel Publishing Co.), p. 145.

Fichtel, C. E., and Trombka, J. I., 1981, *Gamma Ray Astrophysics: New Insight Into the Universe*, NASA SP-453.

Fichtel, C. E., and Kniffen, D. A., 1984, *Astron. Astrophys.*, **134**, 13.

Fichtel, C. E., and Linsley, J., 1986, *Astrophys. J.*, in press.

Galper, A. M. et al., 1975, *15th Internat. Cosmic Ray Conference Papers (Plovdiv)*, **1**, 131.

Georgelin, Y. M., and Georgelin, Y. P., 1976, *Astron. and Astrophys.*, **49**, 57.

Gold, T., 1968, *Nature*, **218**, 731.

Gold, T., 1969, *Nature*, **221**, 25.

Gordon, M. A., and Burton, W. B., 1976, *Astrophys. J.*, **208**, 346.

Graser, U., and Schonfelder, V., 1982, *Astrophys. J.*, **263**, 677.

Grindlay, J. E., 1972, *Astrophys. J.*, **174**, L9-L17.

Grindlay, J. E., 1975, *Astrophys. J.*, **199**, 49.

Grindlay, J. E., 1978, *Nature*, **273**, 211.

Grindlay, J. E., 1982, in *Proceedings of the International Workshop On Very High Energy Gamma Ray Astronomy*, ed. P. V. Ramana Murthy and T. C. Weekes (Ootacamund, India: Tata Institute of Fundamental Research), p. 257.

Gursky, H. et al., 1971, *Astrophys. J.*, **165**, L43.

Harding, A. K., and Stecker, F. W., 1985, *Astrophys. J.*, **291**, 471.

Hartman, R. C. et al., 1979, *Astrophys. J.*, **230**, 597.

Hawking, S. W., 1977, *Scientific American*, **236**, 34.

Hayakawa, S., 1952, *Prog. Theor. Phys.*, **8**, 571.

Hayakawa, S. et al., 1976, *Nature*, **261**, 29.

Helmken, H. F., Grindlay, J. E., and Weekes, T. C., 1975, *14th Internat. Cosmic Ray Conference Papers*, **1**, 123.

Helmken, H. F., and Hoffman, J., 1973, *13th Internat. Cosmic Ray Conference Papers (Denver)*, **1**, 31-35.

Hermsen, W. et al., 1977, *Nature*, **269**, 494.

Hermsen, W., and Bloemen, J. B. G. M., 1982, *Leiden Workshop on Southern Galactic Surveys*.

Hillas, A. M., 1984, *Nature*, **312**, 50.

Horstman, H. M., Cavallo, G., and Moretti-Horstman, E., 1975, *Nuovo Cimento*, **5**, 255.

Houston, B., and Wolfendale, A. W., 1982, *Astron. Astrophys.*, **126**, 22.

Hutchinson, G. W., 1952, *Phil. Mag.*, **43**, 847.

Issa, M. R., Riley, P. A., Strong, A. W., and Wolfendale, A. W., 1981, *J. Phys. G.*, **7**, 656.

Ives, J. C., Sanford, P. W., and Penston, M. V., 1976, *Astrophys. J.*, **207**, L159.

Jelley, T. V., 1966, *Nature*, **211**, 472.

Kafatos, M., and Leiter, D., 1979, *Astrophys. J.*, **229**, 46.

Kanbach, G. et al., 1980, *Astron. Astrophys.*, **90**, 163.

Kazanas, D., and Ellison, D. C., 1986, *Nature,* **319**, 380.

Kifune, T. et al., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **1**, 67.

Kinov et al., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **1**, 67.

Kinzer, R. L., Johnson, W. N., and Kurfess, J. D., 1978, *Astrophys. J.*, **222**, 370.

Kinzer, R. L., Share, G. H., and Seeman, N., 1973, *Astrophys. J.*, **180**, 547-49.

Kinzer, R. L., Share, G. H., and Seeman, N., 1974, *J. Geophys. Res.*, **79**, 4567.

Kniffen, D. A., and Fichtel, C. E., 1970, *Astrophys. J.*, **161**, L157.

Kniffen, D. A., Fichtel, C. E., and Thompson, D. J., 1977, *Astrophys. J.*, **215**, 765.

Kniffen, D. A., Bertsch, D. L., Morris, D. J., Palmeira, R. A. R., and Rao, K. R., 1978, *Astrophys. J.*, **225**, 591.

Kniffen, D. A., Hartman, R. C., Thompson, D. J., and Fichtel, C. E., 1973, *Astrophys. J.*, **186**, L105.

Kniffen, D. A. et al., 1981, *"The Gamma-Ray Observatory Science Plan,"* (NASA).

Komesaroff, M. M., Hamilton, P. A., and Ables, J. G., 1972, *Aust. J. Phys.*, **25**, 759.

Kraushaar, W. L., and Clark, G. W., 1962, *Phys. Rev. Letters*, **8**, 106.

Kraushaar, W. L. et al., 1972, *Astrophys. J.*, **177**, 341.

Lamb, R. C., Fichtel, C. E., Hartman, R. C., Kniffen, D. A., and Thompson, D. J., 1977, *Astrophys. J.*, **211**, L63.

Lamb, R. C., Godfrey, C. P., Wheaton, W. A., and Tümer, O. T., 1982, *Nature*, **296**, 543.

Landecker, T. L., and Wielebinski, R., 1970, *Australian J. Phys. Suppl.*, **16**, 1.

Lebrun, F., 1984, "A Synthetic View at Large Scale of Local Molecular Clouds," *8th European I.A.U., Regional Meeting*, Toulouse, France, September 17-21.

Lebrun, F. et al., 1982, *Astron. Astrophys.*, **107**, 390.

Lebrun, F. et al., 1983, *Astrophys. J.*, **274**, 231.

Leiter, D., and Boldt, E., 1982, *Astrophys. J.*, **260**, 1.

Leiter, D., and Kafatos, M., 1978, *Astrophys. J.*, **226**, 32.

Lloyd-Evans J. et al., 1983, *Nature*, **305**, 784-87.

Maraschi, L., and Treves, A., 1977, *Astrophys. J.*, **218**, L113.

Mayer-Hasselwander, H. A. et al., 1980, *Annals of the New York Academy of Sciences*, **336**, 211-22.

Mayer-Hasselwander, H. A. et al., 1982, *Astron. Astrophy.*, **105**, 164.

McBreen, B. et al., 1973, *Astrophys. J.*, **184**, 571-80.

Meegan, C. A., and Fishman, G. J., 1979, *Astrophys. J.*, **234**, L123.

Meegan, C. A., and Haymes, R. C., 1979, *Astrophys. J.*, **233**, 510.

Montemerle, T., 1979, *Astrophys. J.*, **231**, 95.

Mukanov, J. G. et al., 1979, *16th Internat. Cosmic Ray Conference Papers (Kyoto)*, **1**, 143.

Mushotzky, R. F., 1976, "Observations of Hard X-Rays from Extragalactic Objects." Ph.D. dissertation, University of California, San Diego, 1976.

Mushotzky, R. F., Holt, S. S., and Serlemitsos, P. J., 1978, *Astrophys. J.*, **225**, L115.

Neshpor et al., 1979, *Ap. Space Sci.*, **61**, 349.

Ogelman, H. B., Fichtel, C. E., Kniffen, D. A., and Thompson, D. J., 1976, *Astrophys. J.*, **209**, 584.

Paciesas, W. S., Mushotzky, R. F., and Pelling, R. M., 1977, *Mon. Not. Roy. Astr. Soc.*, **178**, 23.

Page, D. N., and Hawking, S. W., 1976, *Astrophys. J.*, **206**, 1.

Papaliolios, C., and Carleton, N. P., 1971, *The Crab Nebula*, ed. P. D. Davis and E. G. Smith (Dordrecht: D. Reidel Publishing Co., 1977), p. 142.

Parker, E. N., 1966, *Astrophys. J.*, **145**, 811.

Parker, E. N., 1969, *Space Sci. Rev.*, **9**, 654.

Parker, E. N., 1977, in *The Structure and Content of the Galaxy and Galactic Gamma Rays*, NASA CP-002, pp. 283-300.

Parlier, B. et al., 1973, *Nature Phys. Sci.*, **242**, 117-20.

Paul, J. A. et al., 1978, *Astron. Astrophys.*, **63**, L31.

Perotti, F. et al., 1979, *Nature*, **282**, 484.

Perotti, F. et al., 1981, *Astrophys. J.*, **247**, L63.

Pollock, A. M. T. et al., 1981, *Astron. Astrophys.*, **94**, 116.

Pollock, A. M. T. et al., 1985, *COS-B Gamma Ray Sources and Interstellar Gas in the First Galactic Quadrant*, preprint.

Price, R. M., 1974, *Astron. and Astrophys.*, **33**, 33.

Priedhorsky, W., and Terrell, J., 1985, *Long Term Observations of Cyg X-3 with Vela 5B*, preprint.

Primini, F. A. et al., 1979, *Nature*, **278**, 234.

Rankin, J. M. et al., 1970, *Astrophys. J.*, **162**, 707.

Samorski, M., and Stamm, W., 1983, *Astrophys. J.*, **268**, L17.

Schonfelder, V., 1978, *Nature*, **274**, 344.

Schonfelder, V., Graml, F., and Penningsfeld, F. P., 1980, *Astrophys. J.*, **240**, 330.

Schwartz, D.A., and Gursky, H., 1973, *Gamma Ray Astrophysics*, ed. F.W. Stecker and J.T. Trombka, NASA SP-339.

Seiradakis, J., 1976, in *The Structure and Content of the Galaxy and Galactic Gamma Rays*, NASA CO-002, p. 265.

Shapiro, S. L., and Saltpeter, E. E., 1975, *Astrophys. J.*, **198**, 671.

Solomon, P. M., and Sanders, D. B., 1980, in *Molecular Clouds in the Galaxy*, ed. P. M. Solomon and M. Edmunds (Oxford: Pergamon Press), p. 41.

Stecker, F. W., Morgan, D. L., and Bredekamp, J., 1971, *Phys. Rev. Letters*, **27**, 1469.

Stepanian, A. A. et al., 1977, *15th Internat. Cosmic Ray Conference Papers (Plovdiv)*, **1**, 135.

Strong, A. W., and Worrall, D. M., 1976, *J. Phys. A: Math. Gen.*, **9**, 823.

Strong, A. W. et al., 1982, *Astron. Astrophys.*, **115**, 404.

Swanenburg, B. N. et al., 1978, *Nature*, **275**, 298.

Swanenburg, B. N. et al., 1981, *Astrophys. J.*, **243**, L69.

Thaddeus, P., 1982, private communication.

Thompson, D. J., Fichtel, C. E., Kniffen, D. A., and Ogelman, H. B., 1975, *Astrophys. J.*, **200**, L79.

Thompson, D. J., Fichtel, C. E., Kniffen, D. A., and Ogelman, H. B., 1977a, *Astrophys. J.*, **214**, L17.

Thompson, D. J. et al., 1977b, *Astrophys. J.*, **213**, 252.

Thompson, D. J., and Fichtel, C. E., 1982, *Astron. Astrophys.*, **109**, 352-54.

Trombka, J. I., 1977, *Astrophys. J.*, **212**, 925.

Trombka, J. I., Dyer, C. S., Evans, L. G., Bielefeld, M. J., Seltzer, S. M., and Metzger, A. E., 1977, *Astrophys. J.*, **212**, 925.

Tümer, O. T. et al., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **1**, 139.

Vladimirsky, B. M., Stepanian, A. A., and Fomin, V. P., 1973, *13th Internat. Cosmic Ray Conference Papers (Denver)*, **1**, 456.

Wallace, P. T. et al., 1977, *Nature*, **266**, 692.

Webber, W. R., 1983, *Composition and Origin of Cosmic Rays*, ed. M. M. Shapiro (Dordrecht: D. Reidel Publishing Co.), p. 83.

Weekes, T. C., and Helmken, H. F., 1977, *Proc. 12th ESLAR Symp.*, Frascati, Italy (ESA Sp-124), p. 39.

Weekes, T. C., 1978, *Proc. of the 1978 DUMAND Summer Workshop*, 2, ed. A. Roberts (La Jolla: Scripps Institution of Oceanography), 313.

Weekes, T. C., 1983, in *Proceedings of the Workshop on Very High Energy Cosmic Rays, Salt Lake City* (University of Utah).

Weekes, T.C., 1985, "A Third Generation VHE Gamma Ray Observatory," Proceedings of the Workshop on Cosmic Ray & High Energy Gamma Ray Experiments for the Space Station Era, Baton Rouge, Oct. 17-20, 1984, in press.

White, R. S. et al., 1980, *Nature*, **284**, 608.

Worrall, D. M., 1979, *Astrophys. J.*, **232**, 683.

Zyskin, Yu L., and Mukanov, D. B., 1983, *Sov. Astron. Letter.*, **9**, 117.

Zyskin, Yu L., and Mukanov, D. B., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **1**, 177.

N87
24259
UNCLAS

12

NUCLEOSYNTHESIS AND ASTROPHYSICAL GAMMA RAY SPECTROSCOPY

Allan S. Jacobson
Earth and Space Science Division
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, California 91109

1. INTRODUCTION

From its inception, one of the primary motivations for the development of astrophysical gamma ray spectroscopy has been the search for extra-solar system radioactivity. This interest stems from a desire to understand the processes which produced the large array of complex elements that make up our universe and that must have been synthesized subsequent to its creation in the Big Bang. Many of the isotopes produced in such processes are radioactive and their emissions are the key to the detection of their production sites and the identification of the processes involved. How closely the radionuclides and their sources are spatially correlated will, of course, depend upon their decay lifetime and the rate at which they diffuse into the interstellar medium. This topic of investigation was first suggested by Morrison [1958] in his pioneering paper outlining the potential of gamma ray astronomy.

Morrison pointed out that radioactive debris of nucleosynthesis in a supernova might be found in the Crab nebula, the remnant of a supernova explosion seen from Earth in 1054 A.D. Based upon the hypothesis that the decay of ^{254}Cf fueled the exponentially decaying light curve of the supernova [Burbidge et al., 1956] which gave birth to the nebula, he estimated that gamma rays from the decay of ^{226}Ra synthesized in the explosion would be observable today. Savedoff [1959] expanded on these predictions, listing several other radioisotopes which might be detectable if this hypothesis were true. He

estimated that the strongest lines would be at 60 keV from ^{241}Am, and at 180 keV from ^{251}Cf with current fluxes of about 10^{-2} photons/cm^2-s.

Clayton, in a large body of theoretical work [for an example of one such reference, see Clayton, 1982], has firmly established the idea that astrophysical gamma ray spectroscopy is a valuable tool with which to test the theoretical models of nucleosynthesis and to probe the structure of the sites of these high energy processes. He pointed out [Clayton, 1984] that the detection of radioactivity would provide a new cornerstone for theoretical investigations. It was primarily the detailed calculations by Clayton and Craddock [1965] of the gamma rays expected from the Crab nebula under the ^{254}Cf hypothesis which stimulated the earliest flight of a germanium-based high resolution gamma ray spectrometer to search for extra-solar system radioactivity [Jacobson, 1968].

Subsequent theoretical work has shown that the synthesis of ^{254}Cf falls several orders of magnitude short of that required to power the supernova light curve. However, deeper understanding of the processes involved has yielded several other synthesized radioisotopes whose gamma ray line emissions are currently likely candidates for observation. Isotopes produced in Type I supernova are listed in Table 1 [from Lingenfelter and Ramaty, 1978]. The most abundant of these is ^{56}Ni [Clayton, Colgate, and Fishman, 1969]. This isotope decays with a mean life of 8.8 days to ^{56}Co, which has a mean life of 114 days. It decays in turn to ^{56}Fe. Twenty percent of the ^{56}Co decays via positron emission. While the output of gamma ray lines is expected to be intense, a combination of the relatively short lifetimes, dense and absorptive ejecta, and high-expansion velocities with concomitant large Doppler broadening make these lines difficult to observe. Longer lived radionuclides such as ^{26}Al, ^{60}Fe [Clayton, 1974, 1975] and ^{22}Na expected to be produced in abundance with novae [Clayton and Hoyle, 1974] present somewhat greater prospects for observation. Ramaty and Lingenfelter [1977] first noted for reasons discussed below that ^{26}Al would likely be the most detectable of all the radioactive products.

Guided by the list of expected lines in Table 1, a systematic search for cosmic radioactivity was undertaken using the Jet Propulsion Laboratory (JPL) high

Table 1. Gamma ray producing decay chains from explosive nucleosynthesis for an ejecta mass of 1 $M_{\odot}$ [from Lingenfelter and Ramaty, 1978].

DECAY CHAIN	MEAN LIFE (yr)	NUCLEI/ SUPERNOVA	PHOTON ENERGY (MeV)	PHOTONS/ DISINTEGRATION
$Ni^{56} \rightarrow Co^{56} \rightarrow Fe^{56}$	0.31	3×10^{54}	0.847	1
			1.238	0.70
			2.598	0.17
			1.771	0.16
			1.038	0.13
$Co^{57} \rightarrow Fe^{57}$	1.1	7×10^{52}	0.122	0.88
			0.014	0.88
			0.136	0.12
$Na^{22} \rightarrow Ne^{22}$	3.8	3×10^{52}	1.275	1
$Ti^{44} \rightarrow Sc^{44} \rightarrow Ca^{44}$	68	6×10^{51}	1.156	1
			0.078	1
			0.068	1
$Fe^{60} \rightarrow Co^{60} \rightarrow Ni^{60}$	2.2×10^{6}	5×10^{50}	1.332	1
			1.173	1
			0.059	1
$Al^{26} \rightarrow Mg^{26}$	1.1×10^{6}	4×10^{50}	1.809	1

resolution gamma ray spectrometer which flew aboard the HEAO-3 spacecraft. This search resulted in the discovery of ^{26}Al in the interstellar medium [Mahoney et al., 1982; Mahoney et al., 1984].

2. THE HEAO-3 MISSION

The HEAO-3 high resolution gamma ray spectrometer [Mahoney et al., 1980] was the first such instrument to perform an all-sky survey for gamma ray line emissions. It consisted of four approximately 100 cm^3 high purity germanium crystals actively shielded by 6.6 cm thick cesium iodide crystals in

a well-crystal configuration. The aperture was defined by another cesium iodide crystal with a hole drilled through it above each of the Ge crystals, angular collimation of about 30° full width at half maximum (FWHM). A thin plastic scintillator over the aperture provided discrimination against charged particles. Cooling for the Ge crystals for their lifespan of 8.5 months was provided by a two-stage sublimation cooler containing solid methane and ammonia. The instrument made measurements over an energy range of 50 keV to 10 MeV, with an effective area of 72 cm^2 at 100 keV, 27 cm^2 at 500 keV, and 20 cm^2 at 1.5 MeV. A 3-sigma detection sensitivity threshold for a point source was about 2×10^{-4} photons/cm^2-s. The spectral resolution upon initial operation in orbit was 3 keV FWHM at 1.5 MeV.

The HEAO-3 spacecraft was launched on September 20, 1979 and operated until the mission ended on May 30, 1981. The germanium detectors functioned until cryogen exhaustion on June 1, 1980. Subsequent measurement of gamma ray phenomena such as solar flares and gamma ray bursts with the relatively low resolution cesium iodide detectors continued throughout the remainder of the mission. Scanning with a period of 20 minutes, with the spin axis aligned along the sun-earth direction provided a full celestial survey in a six-month period. Additionally, at the beginning of each six-month period, when the solar cell orientation with respect to the Sun made it possible, the spacecraft spin axis was aligned approximately along the galactic spin axis for two weeks, allowing the instrument to scan directly in the galactic equatorial plane.

3. THE HEAO-3 ANALYSIS AND RESULTS

The strategy of looking for radioactive nucleosynthesis materials by concentrating on diffuse galactic emission was suggested by Ramaty and Lingenfelter [1977], in their consideration of supernova ejecta, and by Clayton and Hoyle [1974] in their discussion of nova ejecta. Specifically, they pointed out that isotopes with high production yields and long lifetimes compared with the mean time between the synthesizing events would present the best candidates for detection since they would be observed as diffuse sources which are the cumulative product from a large number of events. In addition, the long lifetime enables the ejecta to slow from its initial velocity to that of the ambient medium before its decay so that the emitted line is not significantly Doppler-

broadened, further enhancing its detectability. Such lines would exhibit broadening only from galactic rotation, which is no more than about 3 keV.

In carrying out such a search, the systematic problems due to background radiation effects are formidable. The radiation background experienced by a spacecraft-borne gamma ray spectrometer in each orbit is complex in both its constituency and its temporal behavior. Cosmic rays modulated by the Earth's magnetic field constantly bombard the spacecraft and its payload, producing a large variety of secondary radiations including gamma ray lines coincident in energy with many of astrophysical importance. Additional radioactivation of the instrument and spacecraft materials and subsequent gamma ray line emission also result from the spacecraft's passage through trapped radiation in the South Atlantic Anomaly (SAA), which occurs several times a day. The primary component of the continuum background results mainly from the modulated cosmic rays and their secondaries and hence is also a function of the magnetic latitude. Auroral X-ray events and clouds of precipitating electrons also contribute their share to the background. These radiation components have characteristic temporal behavior spanning the full range from almost periodic with frequencies on the order of an hour, to transient and exponentially decaying with time constants as small as the order of milliseconds. The analysis is further complicated by a beating between the spacecraft spin period and the orbital period. Accumulating data over many days from a position in the celestial sphere which contains a source of interest, and then comparing that with an equivalent accumulation from a position presumably containing no sources, even when taking as much care as possible that all other parameters such as magnetic latitude, and time since SAA transit, are equal, yields unreliable results. Experience with such difficulties has led to the development of another approach to analyzing data obtained using scanning techniques [Wheaton et al., 1987]. In this approach, each 20-minute scan of the HEAO spacecraft is analyzed independently. Corrections for aperture response are made, and background obtained during the same scan is subtracted. The net result of many scans is then accumulated to give the final net flux.

Since the instrument sensitivity does not allow a measurement of the spatial distribution of a low intensity diffuse source, a model must first be hypothesized and then compared to the data. In analyzing the galactic plane data in

the effort to discover an ^{26}Al component, each scan was fitted with an assumed distribution. While several models have subsequently been used, as discussed below, the initial distribution assumed was that of 100 MeV galactic gamma rays, which is consistent with an extreme population I. The fit included the folding in of instrumental parameters such as detector efficiencies, shield transmission, and the modulation of the source by the Earth. The presence of the Earth was crucial in this approach because it chopped the source emission allowing the background flux level to be established. Because neutron interactions on the ^{27}Al of the spacecraft and instrument produce excited ^{26}Mg, there is a fairly strong instrumental background flux at 1809 keV, the energy of the expected ^{26}Al radiation. The spectral vicinity of the line of interest is shown in Figure 1. It was helpful to the analysis that in close proximity to the target line there was another line at 1778 keV, also produced by neutron interactions on ^{27}Al, and a line at 1764 keV due to the presence of natural ^{238}U contaminants in the instrument materials. A test of the validity of any method used to bring out an astrophysical component of the 1809 keV line is that it must also eliminate these other close lines known to be of background origin.

Figure 2 shows the net result of the analysis. The control emission lines have indeed been removed by the process, leaving a net 1808 keV flux of cosmic origin.

The analysis yields a net line flux of $(4.8 \pm 1.0) \times 10^{-4}$ photons/cm^2-s-rad from the direction of the galactic center. The line has a width of less than 3 keV FWHM, and peaks at an energy of 1808.49 ± 0.41 keV. The accepted energy of the gamma ray line due to the decay of ^{26}Al is 1808.65 ± 0.07 keV.

An effort has been made to better understand the spatial distribution of the gamma ray emission. This is vital to any determination of its source. Distributions other than that of the high energy gamma rays have been fit to the data [Mahoney et al., 1985]. These are: (1) the extreme population I, containing the type II supernovae and massive main sequence stars, believed to be represented by the galactic CO distribution [Burton and Gordon, 1978] and (2) the total galactic visual luminosity, [Bahcall and Soneira, 1980] believed representative of novae [Ciardullo, 1984] and red giants. Both of these distributions fit the observations reasonably well, and cannot be distinguished with

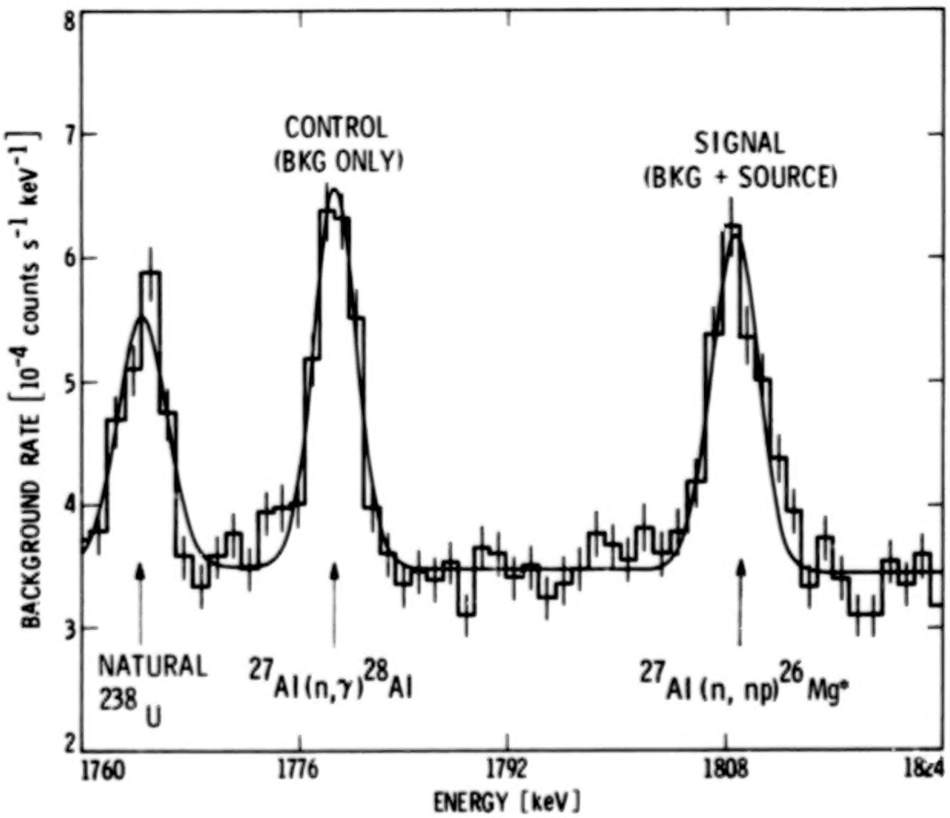


Figure 1. The background spectrum in the vicinity of the 1809 keV line. These lines are traceable in the main to either natural radioactive impurities, or emissions resulting from interactions by cosmic ray produced neutrons with the instrument and spacecraft materials.

the data. A brief study has been made of other source distributions. However, the statistical significance of these fits is such that the only distributions which can reasonably be ruled out are those which are very highly peaked toward the galactic center with full width at half maximum of less than about 6 degrees. The flux value from the direction of the galactic center depends upon the

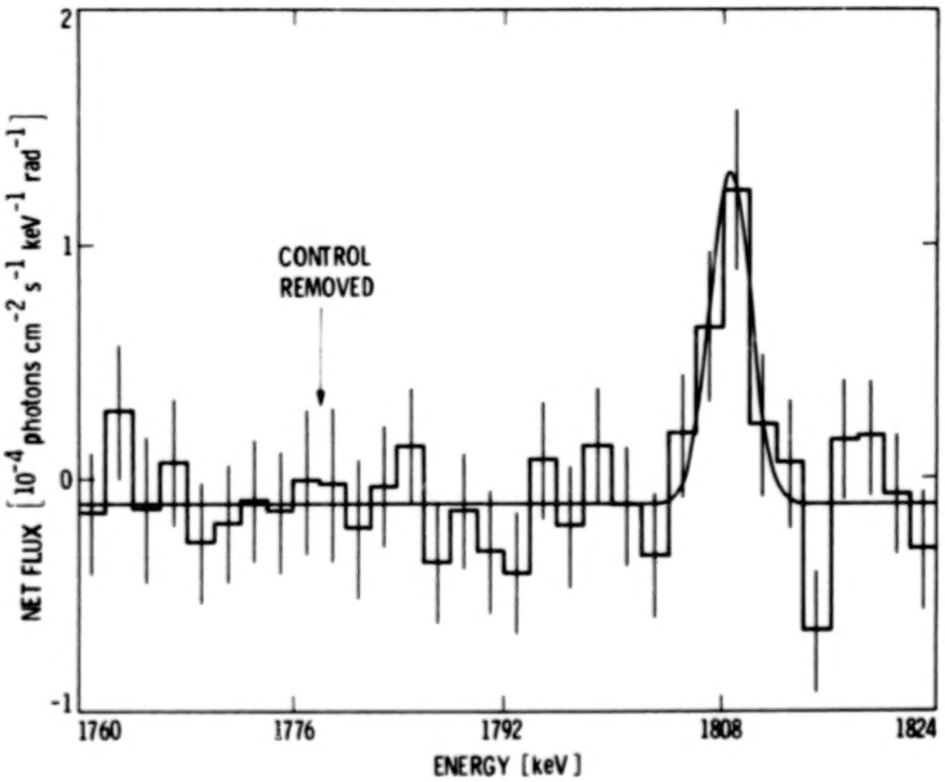


Figure 2. The net diffuse galactic gamma ray flux near 1809 keV. The process of analysis has eliminated the background lines and revealed the net cosmic component of the 1809 keV line.

distribution chosen, but for distributions which are statistically acceptable, the statistical significance of the line varies little.

In order to determine whether the peak of the distribution of ^{26}Al gamma rays is in the direction of the galactic center or some other direction indicating perhaps a local origin, fits of the extreme population I distribution were made to the data assuming the centroid of the distribution to be successively at longitudes spaced by 60 degrees around the galactic disk. Figure 3 shows the results of these fits where the statistical significance of the net flux from the

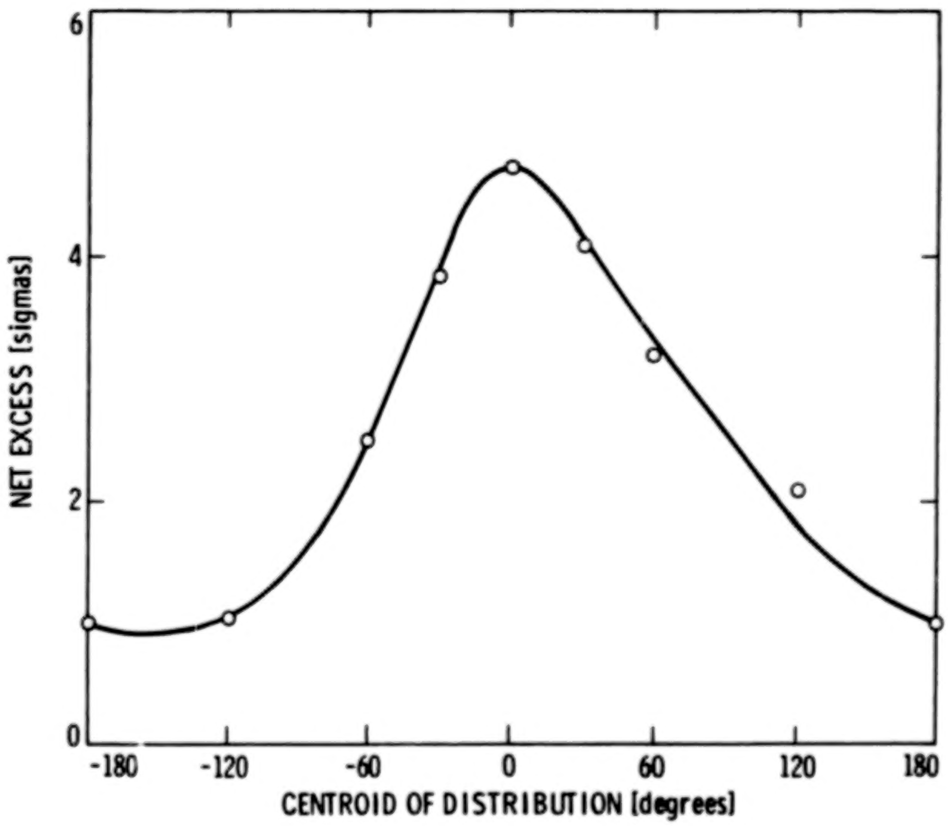


Figure 3. Significance of the 1809 keV observation as a function of galactic longitude.

direction of the centroid of the distribution is plotted against the centroid position. A similar study was carried out in galactic latitude. The net results of these studies shows that the emission is centered at $\ell = -6 \pm 22$ degrees and $b = -4 \pm 20$ degrees.

Share et al. [1985] have reported a confirming observation of the ^{26}Al galactic emission using the gamma ray detector aboard the Solar Maximum Mission satellite. In an analysis of 3.5 years of data, a net source of gamma ray line radiation was detected at an energy of 1804 ± 4 keV. Assuming a source

distribution like that of the high energy gamma rays, they found a net flux from the direction of the galactic center of $(4.0 \pm 0.4) \times 10^{-4}$ photons/cm^2-s-rad. The peak of the distribution was in an error box (99% confidence) defined by 345 and 25 degrees in galactic longitude and -15 and $+10$ degrees in latitude. These results agree in all respects with those of the HEAO-3 observations.

4. THE SOURCE OF THE ^{26}Al

An estimate of the total mass of ^{26}Al presently distributed in the galaxy can be arrived at in the following manner. It has been shown by Higdon and Lingenfelter [1976] that for a diffuse galactic gamma ray source with an extreme population I distribution, the measured flux, F, from the vicinity of the galactic center can be related to the total galactic luminosity, Q, by

$$F \text{ (photons/cm}^2\text{-s-rad)} = 1 \times 10^{-46}\ Q \text{ (photons/s)}.$$

Given the measured flux from the galactic center direction of $F = 4.8 \times 10^{-4}$ photons/cm^2-s-rad, it is found that the galactic luminosity in 1809 keV gamma rays is $Q = 4.8 \times 10^{42}$ photons/s. For a ^{26}Al mean lifetime of 1.04 $\times 10^6$ years, the total mass of ^{26}Al presently in the galaxy is then about 3 M_O. A similar value for the mass of ^{26}Al is found under the assumption that the distribution follows the visual distribution model [Mahoney et al., 1985]. By taking the mass of the galactic interstellar medium to be 4×10^9 M_O [Salpeter, 1977] and the present average galactic mass fraction of ^{27}Al to be 6.6×10^5 [Cameron, 1982], we find that there is approximately 2.6×10^5 M_O of ^{27}Al in the galaxy. The present average galactic ration of ^{26}Al/^{27}Al is therefore about 1×10^{-5}.

Current models [Woosley and Weaver, 1980] indicate that supernova production is too low by at least an order of magnitude to explain the observed ^{26}Al [Clayton, 1984]. In novae, the production ratio of ^{26}Al/^{27}Al is approximately unity [Hillebrandt and Thielemann, 1982; Clayton, 1984], so that for a nova rate of 40 per year, an average ejected mass of about 10^{-4} M_O, and a ^{26}Al mass fraction in the ejecta of 2.6×10^{-4}, one can expect that there is about one solar mass of ^{26}Al from this process currently in the galaxy. This

is the right order to account for the HEAO-3 observation. Hillebrandt and Thielemann [1982] have pointed out that the high productivity of novae in synthesizing ^{26}Al is rather insensitive to details of the nova model, but intimately reflects the properties of the nuclear reactions involved. Recent new results on the ^{26}Al production cross-sections will perhaps cause these estimates to be revised upward. Champaigne, Howard, and Parker [1983] have found one and perhaps two resonances in the cross section for the hydrogen burning reaction $^{25}Mg\ (p,\gamma)^{26}Al$, the main production reaction for ^{26}Al. These resonances tend to greatly increase the reaction rate at lower temperatures. These results require a reevaluation of the nova yields and open the possibilities for other stellar sources of ^{26}Al.

Norgaard [1980] has pointed out that significant amounts of ^{26}Al can be made at the base of the outer convective envelope in a red giant star. Cameron [1984] estimates that based upon the revised rate for $^{25}Mg\ (p,\gamma)\ ^{26}Al$, the case for this has been greatly strengthened and that a considerable fraction of the ^{25}Mg in the envelope of red giant stars may be converted to ^{26}Al. Pending a more precise calculation, he estimates that $^{26}Al/^{27}Al$ ratios of from unity to as much as 10 could occur and that if 0.1% or more of the interstellar medium has been recycled through red giant stars, then it is possible that the observed ^{26}Al may be principally contributed by them.

Blake and Dearborn [1984] and Dearborn and Blake [1985], have pointed out that type O and Wolf-Rayet stars will produce substantial amounts of ^{26}Al, and disperse the material in their intense stellar winds. They estimate that 0.5 M_O of ^{26}Al in the interstellar medium is traceable to this source.

In addition to the provocative theoretical work on the problems of nucleosynthesis, much interest in the production of ^{26}Al has been stimulated by the discovery of its decay products in primitive solar system materials. The study of select inclusions in meteorites lead to conclusions that a significant amount of radioactive ^{26}Al was present in the solar nebula when condensation took place.

The discovery of anomalously high quantities of ^{26}Mg, the decay product of ^{26}Al in primitive solar system materials [Lee, Papanastassiou, and Wasserburg, 1977] provided the most important reason leading to speculation that

the formation of the solar system was triggered by a supernova close enough to the protosolar nebula to induce it to collapse as well as inject ^{26}Al [Cameron and Truran, 1977]. As pointed out by Cameron [1984], none of the arguments previously advanced for this model seems very compelling today. The HEAO-3 discovery of relatively large amounts of ^{26}Al in the interstellar medium suggests that the meteoritic inclusions reflect either normal interstellar concentrations or a nearby production site of rather more common occurrence than a supernova.

5. FUTURE EXPERIMENTS

The statistical limitations on the observations made thus far leave quite a number of unanswered questions about the distribution of ^{26}Al, and the source or sources of this material. While theoretical work can provide suggestions for specific measurements, these questions can ultimately be addressed only by further observations. The spatial distribution must be mapped in detail both in galactic longitude and latitude. Even angular resolution of about 5 $\times$ 5 degrees and a flux sensitivity five times greater than that of the HEAO-3 experiment would allow discrimination between extreme population I, and an older disk population.

High spectral resolution will also be significant in determining the spatial distribution of ^{26}Al through the study of velocity dispersions. Energy resolution presently achievable with germanium detectors would allow measurements of Doppler broadening of about 1 keV FWHM, permitting study of velocity dispersions as small as ± 85 km/s. For instance, 30 percent of galactic intermediate population stars such as novae and red giants are in the galactic ellipsoidal bulge, concentrated toward the galactic nucleus [Higdon, 1985]. This population has a velocity dispersion of about 130 km/s. On the other hand, extreme population I stars concentrated in the disk have a velocity dispersion of about 30 km/s. Independent of angular resolution, improved spectral resolution and sensitivity will separate the galactic bulge and disk components of ^{26}Al. The portion of the ^{26}Al still exhibiting velocities associated

with its production sites ($10^3 - 10^4$ km/s) and dispersion processes may also be observable.

The observations need to be extended to other products of nucleosynthesis. The immediate candidates are those shown in Table 1. The measurement of a galactic 511 keV line would serve as an indicator of explosive ^{56}Fe synthesis in type I supernovae. Positrons are produced in the decay of ^{56}Co and a fraction (perhaps 10%) are believed to escape from the expanding shell and annihilate in the interstellar medium over 10^5 to 10^6 years. Thus one could see a narrow line with an intensity comparable to the 1809 keV line. Its spatial distribution could also have important implications for the propagation and acceleration of cosmic rays. Analysis of this component is underway at present in the HEAO-3 data. The measurement of the decay lines of ^{60}Fe would demonstrate ongoing nucleosynthesis in massive young stars via helium burning in type II supernovae. The lines are also expected to be narrow with an intensity of order 10% that of the 1809 keV line. Measurement of the spatial distribution would indicate regions of massive star formation in the galaxy. The isotope ^{44}Ti is produced in type I supernovae (SN I). The measurement of line emission from this isotope would identify recent (of order 100 years) and previously known SN I remnants in the galaxy.

For the reasons given above, high resolution spectroscopy is an important part of any gamma ray astronomy program. At this time, however, there is no high resolution spectroscopy experiment currently planned for a space mission. There was originally a high resolution gamma ray spectrometer as part of the payload of the Gamma Ray Observatory (GRO) to be launched in 1988. For budgetary reasons, NASA decided to remove this experiment from the complement of instruments, rationalizing that the low resolution scintillation spectrometer aboard the mission could satisfy the spectroscopic objectives of low energy gamma ray astronomy. This design was made before many of the current discoveries in this spectral region. It was based in part on an expectation that narrow gamma ray lines would not be seen. In fact, at that time the only confirmed and well established extra-solar system line, the galactic center .511 MeV line *was* narrow. With the possible exception of lines in gamma ray bursts and solar flares, only narrow lines have been observed thus far. The width of these lines is crucial to the understanding of their sources. The narrow width ($\lesssim$ 3 keV) of the galactic center line sets important limits on

the gravitational potential, temperature, and degree of ionization of the positron-electron annihilation medium [Leventhal, MacCallum, and Stang, 1978; Riegler et al., 1981; Riegler et al., 1985]. Thus, the failure to include a high spectral resolution instrument in the present program was, in retrospect, unfortunate.

Scintillation spectrometers such as the one presently in the GRO payload, with spectral resolution some 25 or more times worse than germanium spectrometers, have made valuable contributions to the field of gamma ray spectroscopy, notably in the areas of solar flare [see Chupp, 1982 and Chupp, 1984] and gamma ray burst [Mazets et al., 1981] measurements. In these observations, the fluxes have been large compared to the detector background, and the duration of the observations has been short so that systematic effects become less important. However, even in these types of observations, the relatively poor spectral resolution has limited the degree to which line width measurements can characterize the emission regions. Additionally, if the spectra are highly complex, as in the case of some of the solar flares, the identification of the components becomes difficult or impossible.

There are other factors beside the spectral resolution which seriously limit the ability of scintillation detectors to make long duration observations under low signal-to-background conditions, regardless of whether narrow or broad gamma ray lines are expected. Both types of instruments, their structural materials, and their vehicles are subject to cosmic ray bombardment which initiates nuclear reactions. These reactions result in numerous gamma ray lines in the background. Figure 4 shows the background as measured by the HEAO-3 high resolution gamma ray spectrometer. There are more than 140 lines, each exhibiting characteristic time variation as the spacecraft executes its various orbital motions, passages through the South Atlantic Anomaly, or continues to be exposed to the ambient radiation environment. While the details of which lines are present, and what their intensities and time variations may be, both germanium and scintillation detectors, and any other materials used in these instruments have backgrounds of similar complexity.

The ability to deal with this systematic background and compensate for it depends upon the experimenter's ability to detect, identify, and characterize each component. The high resolution of germanium allows identification of

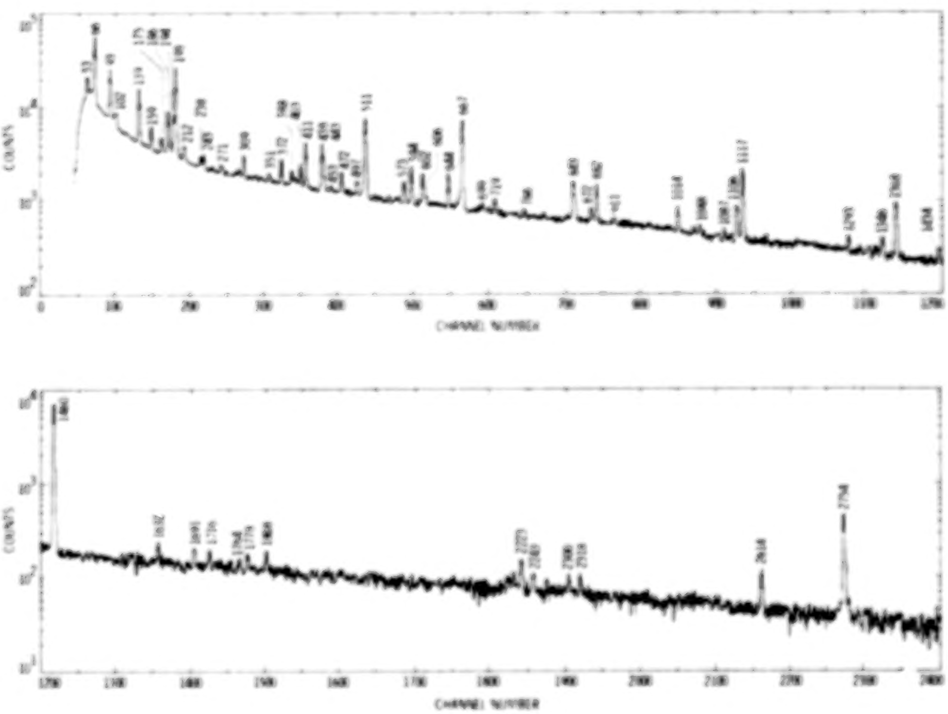


Figure 4. A background spectrum measured in one of the HEAO-3 detectors. The spectrum represents a four-day accumulation.

most of the lines and hence the characteristic of their sources. With scintillation detectors, such identifications are largely guess work. This introduces large systematic errors which are extremely difficult to deal with and hence often overlooked when reporting the results of measurements.

The history of non-solar gamma ray spectroscopy with scintillators also suggests a cautious attitude about the expectations of the GRO low energy gamma ray detector to make low intensity spectroscopic measurements. Several expeditions to the Southern Hemisphere to observe the galactic center with a scintillation spectrometer did yield positive measurements of a line near .511 MeV [Johnson and Haymes, 1973; Haymes et al., 1975] but never close enough

to the energy .511 MeV to be convincingly identified as positron-electron annihilation radiation. Not until observations with germanium spectrometers began [Leventhal, MacCallum, and Stang, 1978] was it established that in the direction of the galactic center there was a source of positron-electron annihilation radiation.

A situation paralleling the original GRO project existed in the HEAO program, which included *both* scintillation and germanium spectrometers. Reviewing the results obtained by these experiments to date, it is striking how little overlap there has been. In particular, essentially none of the narrow line results from the HEAO high resolution spectrometer have been accessible to the HEAO-1 scintillation spectrometer. The reason is that NaI and Ge spectrometers have very different capabilities, making them complementary rather than competitive. Because NaI is cheaper than high purity Ge, and does not require cooling to cryogenic temperatures, a scintillation experiment will be cheaper and simpler per square centimeter of effective area than a high resolution spectrometer of similar proportions. But scintillators simply cannot address whole areas of astronomical observation which are accessible to Ge. It is extremely difficult in practice for scintillation instruments to convincingly demonstrate the existence of weak lines, as is shown by the HEAO program experience with the 511 keV and 1809 keV lines. When one move beyond mere existence to ask questions about the energy, width, and profile of these lines, in which a great richness of astrophysical information is contained, "extremely difficult in practice" becomes "obviously impossible". Because there is no high resolution spectrometry experiment on the GRO, such questions will probably not be answered in the next five years.

To do the necessary and comprehensive job of observationally establishing the source of the ^{26}Al and to extend these observations to other products and processes of nucleosynthesis requires the development and space flight of a high resolution gamma ray spectrometer. Specifically, space flight rather than balloon flights are needed because the latter, while providing a somewhat more benign environment with regard to sources of systematic error, drastically limit: (1) the number of sources available in a reasonable time for study, (2) the observing time achievable upon any single source, and (3) the studies which can be made of large-scale diffuse sources. An impractically large number of balloon flights would be required to adequately address these subjects. It

is a sobering thought that the usable observing time of the HEAO-3 gamma ray spectrometer was longer than the combined observing time of all the gamma ray astronomy balloon flights ever made.

Recent missions have tended to be large and consequently expensive observatories. Because of the expense, the frequency of mission opportunities has been reduced. The likelihood of on-orbit servicing and repair of satellite payloads could extend the lifetimes of these observatories and contribute further to the scarcity of opportunities to launch a large and definitive high resolution spectrometer, so that a decade or more could elapse before the important measurements discussed above are carried out. I believe that every effort should be made by interested scientists and by NASA to see that an opportunity to fly a high resolution gamma ray spectrometer occurs in the near future.

6. SUMMARY

The HEAO-3 gamma ray spectrometer has provided new evidence in the quest for the understanding of complex element formation in the universe with the discovery of ^{26}Al in the interstellar medium. It has demonstrated that the synthesis of intermediate mass nuclei is currently going on in the galaxy. This discovery has been confirmed by the Solar Maximum Mission. The flux is peaked near the galactic center and indicates about 3 $M_\odot$ of ^{26}Al in the interstellar medium, with an implied ratio of $^{26}\mathrm{Al}/^{27}\mathrm{Al} = 1 \times 10^{-5}$. Several possible distributions have been studied but the data gathered thus far do not allow discrimination between them. It is felt that only the spaceflight of a high resolution gamma ray spectrometer with adequate sensitivity will ultimately resolve the issue of the source of this material.

The author wishes to acknowledge valuable suggestions by his colleagues, J. C. Higdon, J. C. Ling, W. A. Mahoney, R. A. Schwartz, and Wm. A. Wheaton.

REFERENCES

Bahcall, J. N., and Soneira, R. M., 1980, *Astrophys. J.* (Suppl.), **44**, 73.

Blake, J. B., and Dearborn, D. S. P., 1984, *Adv. Space Res.*, **4**, 89.

Burbidge, G. R., Hoyle, F., Burbidge, E. M., Christy, R. F., and Fowler, W. A., 1956, *Phys. Rev.*, **103**, 1145.

Burton, W. B., and Gordon, M. A., 1978, *Astrophys. J.*, **63**, 7.

Cameron, A. G. W., 1982, in *Essays in Nuclear Astrophysics*, ed. C. A. Barnes, D. D. Clayton, and D. N. Schramm (Cambridge University Press), p. 23.

Cameron, A. G. W., 1984, *Icarus*, **60**, 416.

Cameron, A. G. W., and Truran, J. W., 1977, *Icarus*, **30**, 447.

Champaigne, A. E., Howard, A. J., and Parker, P. D., 1983, *Astrophys. J.*, **269**, 686.

Chupp, E. L., 1982, in *Gamma-Ray Transients and Related Astrophysical Phenomena*, ed. R. E. Lingenfelter, H. S. Hudson, and D. M. Worrall (New York: A.I.P.), p. 363.

Chupp, E. L., 1984, *Ann. Review of Astron. and Astrophysics*, **22**, 359.

Ciardullo, R., 1984, *BAAS*, **16**, 977 and 1985, private communication.

Clayton, D. D., 1974, *Astrophys. J.*, **188**, 155.

Clayton, D. D., 1975, *Astrophys. J.*, **198**, 151.

Clayton, D. D., 1982, *J. R. Astr. Soc.*, **23**, 1974.

Clayton, D. D., 1984, *Astrophys. J.*, **280**, 144.

Clayton, D. D., Colgate, S. A., and Fishman, G. J., 1969, *Astrophys. J.*, **155**, 75.

Clayton, D. D., and Craddock, W. L., 1965, *Astrophys. J.*, **142**, 189.

Clayton, D. D., and Hoyle, F., 1974, *Astrophys. J.*, **187**, L101.

Clayton, D. D., and Hoyle, F., 1976, *Astrophys. J.*, **203**, 490.

Dearborn, D. S., and Blake, J. B., 1985, *Astrophys. J.*, **288**, L21.

Haymes, R. C., Walraven, G. D., Meegan, C. A., Hall, R. D., Djuth, F. T., and Shelton, D. M., 1975, *Astrophys. J.*, **201**, 593.

Higdon, J. C., 1985, private communication.

Higdon, J. C., and Lingenfelter, R. E., 1976, *Astrophys. J.*, **208**, L107.

Hillebrandt, W., and Thielemann, F. K., 1982, *Astrophys. J.*, **255**, 617.

Jacobson, A. S. "A Search for Gamma-Ray Line Emissions from the Crab Nebula." Ph.D. dissertation, University of California, San Diego, 1968.

Johnson, W. M., and Haymes, R. C., 1973, *Astrophys. J.*, **184**, 203.

Lee, T., Papanastassiou, D. A., and Wasserburg, G. J., 1977, *Astrophys. J.*, **221**, L107.

Leventhal, M., MacCallum, C. J., and Stang, P. D., 1978, *Astrophys. J.*, **225**, L1.

Lingenfelter, R. E., and Ramaty, R., 1978, *Physics Today*, **31**, 40.

Mahoney, W. A., Ling, J. C., Jacobson, A. S., and Lingenfelter, R. E., 1982, *Astrophys. J.*, **262**, 742.

Mahoney, W. A., Ling, J. C., Jacobson, A. S., and Tapphorn, R. N., 1980, *Nucl. Inst. & Methods*, **178**, 363.

Mahoney, W. A., Higdon, J. C., Ling, J. C., Wheaton, W. A., and Jacobson, A. S., 1985, *19th Internat. Cosmic Ray Conference Papers (La Jolla)*, **1**, 357.

Mahoney, W. A., Ling, J. C., Wheaton, W. A., and Jacobson, A. S., 1984, *Astrophys. J.*, **286**, 578.

Mazets, E. P., Golenetskii, S. V., Aptekar, R. L., Gur'yan, Y. A., and Il'inskii, V. N., 1981, *Nature*, **290**, 378.

Morrison, P., 1958, *Il Nuovo Cimento*, **7**, 858.

Norgaard, H., 1980, *Astrophys. J.*, **236**, 895.

Ramaty, R., and Lingenfelter, R. E., 1977, *Astrophys. J.*, **213**, L5.

Riegler, G. R., Ling, J. C., Mahoney, W. A., Wheaton, W. A., and Jacobson, A. S., 1985, *Astrophys. J.*, **294**, L13.

Riegler, G. R., Ling, J. C., Mahoney, W. A., Wheaton, W. A., Willett, J. B., and Jacobson, A. S., 1981, *Astrophys. J.*, **248**, L13.

Salpeter, E. E., 1977, *Ann. Rev. Astr. Ap.*, **15**, 267.

Savedoff, M. P., 1959, *Il Nuovo Cimento*, **13**, 1584.

Share, G. H., Kinzer, R. L., Kurfess, J. D., Forrest, D. J., Chupp, E. L., and Rieger, E., 1985, *Astrophys. J.*, **292**, L61.

Wheaton, W. A., Ling, J. C., Mahoney, W. A., and Jacobson, A. S., 1987, to be submitted to *Astrophys. J.*

Woosley, S. E., and Weaver, T. A., 1980, *Astrophys. J.*, **238**, 1017.

N87
24260
UNCLAS

N87-24260

13

GAMMA RAY TRANSIENTS

Thomas L. Cline
Laboratory for High Energy Astrophysics
Goddard Space Flight Center
Greenbelt, Maryland 20771

1. PROLOGUE

History has had its periods of magnificent excess, when vast amounts of human effort were devoted to cultural projects that transcended the necessities of survival and commerce. Obvious examples begin with the construction of the great pyramids. Much later, Europe's cathedrals were the ultimate commitment of society's surplus energy to expressions of the spirit. The explorations of the remote regions of the Earth followed in time as luxuries of civilization that were also adventurous investments. During the last century, technologies of all kinds were obsessively undertaken, with results that have transformed life. Most recently, and for a brief moment fitting onto this logarithmically shrinking temporal series, were the years of the intense involvement of our segment of society in the exploration of space. This endeavor is generally thought of as having things in common with each of those earlier enterprises; it might even be considered as an evolutionary culmination of their entire trend.

The primary motivations for the American space effort during its first few years were the immediate and compelling issues of national image and defense. The civilian space agency was the most visible response to the Sputnik challenge, but not the country's principal space program, becoming second or third, depending on the bookkeeping, behind both the defense space development and the surveillance budgets. The desire to fulfill its mission of promoting the national image ensured that NASA emphasize manned exploration—rather than focusing only on remote sensing with automated instrumentation—in order to enlist and maintain the enthusiastic support of the public. As the technologies proved themselves with the early successes,

and as the race to the Moon was won, practical enterprises such as navigation, communication, and weather monitoring grew in scope. The large-scale lunar and deep space activities have given way to the shuttle and, more recently, to the space station, near-earth projects intended to prepare for the systematic future growth into space.

My point is that scientific research was never a necessary ingredient in any of this. Appealing close-up photos of the distant objects of the solar system notwithstanding, pure research for its own sake was, as always, a luxury of which the typical spectator, or voter, was almost completely unaware. I like to think that Frank McDonald was more responsible than any other one person for the way that science, in fact, did figure into the space arena.

2. INTRODUCTION

The diversity of the accounts that might attempt to describe Frank McDonald's influence on the course of space science would be considerable, no doubt scattering throughout a diagram in argument space of at least four dimensions. My opinions are therefore entirely my own. In the early days there were, of course, a number of creative individuals who were deeply committed to scientific excellence as the first priority. These were the people who promoted space science as a valuable function for Goddard as a federal center, or who sought to constructively influence NASA program creation, or who administered research empires, or who also valued the maintenance of scientific activities as a support service to NASA, citing solar protons as health hazards in space, for example. McDonald is one individual who seemed to combine all of these career objectives simultaneously, even while pursuing personal research of considerable merit and coaching apprentices and thesis students and teaching on the side.

Entirely in Frank's own style, as well, was the generation of new research capabilities. Nurtured within his lab, these groups worked in areas generally outside his own specialty of cosmic radiation although potentially related with some interdisciplinary values. Following various growth rates, and either transplanted elsewhere or remaining, these have come to range from mature

individual scientists sharing support facilities to competitive laboratory empires. A creation of this nature is a complex, simultaneously insurance-buying, sphere of influence-extending, and yet selfless, risk-taking, and entirely existential act. On a comparatively minor scale, even McDonald's work-a-day ideas and advice could result in significant career opportunities or alterations. One such episode in the early 1970's made possible a unique development in astrophysics, in my opinion, and is what I wish to use as illustration here.

3. HISTORY

The discovery of cosmic gamma ray bursts was published after more than a dozen brief and inexplicable increases of the 100 keV count rate in cislunar space had been observed in the course of a nuclear test-band monitoring program [Klebesadal, Strong, and Olson, 1973]. This serendipitous space age discovery was made, not with NASA or European scientific instrumentation, but with systems designed at the Los Alamos Laboratory for the detection of nuclear explosions beyond the atmosphere. Its release was quite conservatively delayed until after several years of confidence building through the consistent accumulation of data. Analysis of the onset times of each rate increase at the widely separated orbiting instruments, not exactly in coincidence, could geometrically define event propagation planes. The pattern of the source directions found in this manner was consistent with isotropy and bore no relation to the locations of the Earth, Moon, or Sun. The gamma ray counting rates were similar to those from solar flares, obviously indicating much greater total emission, at least billions of times greater, if indeed coming from outer space. Since the events were not found to be from supernovae, the only mechanisms then suggested as sources of observable cosmic gamma ray transients [Colgate, 1968], and since their discovery predated that of X-ray bursts, gamma ray bursts were a real surprise.

New life was thus injected both into gamma ray astrophysics and into the astronomy of transients. Despite an earlier prediction that exotic forms of cosmic information might be transmitted in nuclear gamma rays [Morrison, 1985], this energy region had continued to be disappointing. Compared with the X-ray band, that has been more than generous in its rewards in astronomy,

the 0.1 to 10 MeV region then seemed to require such elaborate instrumentation for equivalent signal strengths as to be almost hopelessly impractical. Yet, here was an unexpected phenomenon with counting rates practically off scale, found with very small and relatively crude detectors, and having energy releases certain to be enormous, even though the sources were as yet unidentified. Further, the curiosity value of this discovery was enhanced by the fact that these observations (in this very wavelength region hitherto encompassing little more than a static form of neutral cosmic ray background) were characterized by almost immeasurably rapid time scales. The excitement of the moment, while prompting many imaginative ideas and speculations, left experimenters unable to immediately conduct new observations in response, given that the detection rate was far too low for a rocket or balloon flight to have any chance of being aloft during an event and that spacecraft experiments took years to get into the mission schedules and to get done.

At that time I had just started a low energy gamma ray astronomy effort at Goddard. The primary objective, coincidentally enough in retrospect, was gamma ray transients, then thought to be possibly observable as signatures of distant supernovae [Colgate, 1968]. I was building a balloon payload for this purpose, as well as searching through and comparing existing spacecraft data records for evidence of transient gamma ray behavior. Only weeks before the release to the public of the discovery of bursts by Los Alamos, several anomalous counting rate increases had been found in my IMP-eye solar flare hard X-ray data. These were detected outside the Earth's environment and were also apparently of non-solar origin. We could find independent evidence for only one of them in other spacecraft records. That was, however, not very firm evidence, being a single time coincidence in the list of hundreds of fluctuations that had been accumulated by an OSO-7 instrument in near-earth orbit. When the Los Alamos list was published, our IMP-eye bursts, all of which appeared on that list, both gave immediate confirmation to the gamma ray burst phenomenon and included the first burst energy spectrum, demonstrating its distinctly non-X-ray nature [Cline et al., 1973].

With the single OSO-IMP event, a more extended spectrum was obtained, together with the first approximate source direction confirmation [Wheaton et al., 1973]. The latter was also of particular interest at that time, due to the uncertainties inherent in Los Alamos' technique of finding the burst

wavefront vector using the relative Vela satellite burst 'trigger' timings. This single OSO source field, although quite large, was typical of the existing pattern, having an anomalous direction at high galactic latitude, far from the brightest X-ray emitters, that, of course, enhanced the source mystery and provided even more confirmation that an entirely prime scientific phenomenon was ripe for exploration.

With Frank McDonald's appreciation of this opportunity and with his administrative and moral support, I was able to believe that I had some chance of demonstrating the urgency of the situation to NASA Headquarters. I well remember promoting—in the company of other interested persons, including George Pieper, Les Meredith, and Doyle Evans—the opportunities that then existed, which, if ignored or unnecessarily postponed, could be left wide open to other space groups or missed entirely. John Naugle, to his great credit, endorsed the two propositions that physics and astronomy instruments then under construction not be excluded from the opportunity for modifications to include gamma ray burst observational capabilities, if appropriate and at low cost, and that competitive proposals for new experiments, even on space missions not necessarily devoted to astronomy, be considered for possible gamma ray burst instruments, when appropriate and at minimal cost. This action made possible the American participation in the gamma ray burst interplanetary network. In retrospect, it probably enabled all the domestic spaceborne gamma ray burst studies to be carried out in the entire one and a half decades before Gamma Ray Observatory, with the exception of Solar Maximum Mission. The situation otherwise would not have been a multivertex network capable of high accuracy 'triangulation,' but a single long baseline from several near-earth instruments to the various Franco-Soviet instruments on the Veneras.

The experiences of those days began a commitment that continues to the present. While developing new ways to research this puzzle, our previously scoped gamma ray transient studies were redirected towards gamma ray bursts. The inconclusive results of our first balloon-borne experiment prompted us to put up in the following year two balloon instruments simultaneously, with a distance separation of several midwestern states. This search for smaller sized and hopefully more frequently occurring events incorporated the obvious requirement that independent detections would be needed to establish an

ephemeral effect as real. The results [Cline et al., 1976] were like those that plague even present-day high sensitivity balloon searches [e.g., Meegan, Fishman, and Wilson, 1985], namely, a lack of detection of the weak bursts that should seem to be required from a reasonably extended size distribution.

4. OBSERVATIONS

Helios-2 was the first gamma ray burst instrument launched; its initial results, in 1976, seemed to deepen the mystery. The great distance of this solar orbiter, of up to two astronomical units, made possible the determination of considerably more definitive source location loci. The comparison of its measurements with those from the near-earth Vela system provided source fields in the form of ring segments as narrow as a minute of arc, although up to tens of degrees in length. Sources were not, of course, identifiable with such observations, but candidate sources could be unambiguously eliminated by the lack of positional agreement. All the burst observations showed a clear and complete lack of agreement with the locations of all obvious candidate sources such as the well-known X-ray objects [Cline et al., 1979]. This problem of source identification (at least for the 'classical' types of >100 keV character, as discussed below) continues in some form to the present.

The first interplanetary burst network was completed in 1978 with the launches of Pioneer Venus Orbiter and Veneras-11 and -12. A variety of instruments on these spacecraft, flown to and beyond the planet Venus, provided the necessary third vertex in the 'triangulation' array, complementing those near the Earth and Helios-2 in its solar orbit. Later, the Veneras were to outdistance Venus, giving a multiply determined array. Also, the several burst detectors piggybacked on the third International Sun Earth Explorer (ISEE) supplied considerable improvement in the accuracy of the near-earth data useful for this network. With its initial results, this progression of gamma ray burst source directions determined with ever-increasing accuracy (that turned out not to be consistent with the positions of known objects) reached its conclusion. Source fields were derived with sizes from tens of arc-seconds to arc-minutes in dimension, yielding precise 'error boxes' sufficiently small to establish source identification, if reasonable agreement were to be found, but (with the particular exception noted below) finding fields that were either entirely empty

or empty of other than random stellar foreground objects [Laros et al., 1981, illustrated in Figure 1; Cline et al., 1981; Barat et al., 1984a, b; and Cline et al., 1984]: Clearly, not only were the gamma ray burst sources extremely elusive, but any of their companions that may exist in binary associations as well.

One totally unexpected discovery made possible by these results was that of an archived optical transient [Schaefer, 1983], precisely within the November

Figure 1. One of the initial gamma ray burst source fields determined with the interplanetary network [Laros et al., 1981], typical in its absence of apparent optical counterparts. These also have no X-ray counterparts.

19, 1978 burst source field [Cline et al., 1984]. Figure 2 shows comparison photos. The flash, found in a search through hundreds of stored prints, has somewhat of a 'supernova' appearance, but in fact lasted less than several minutes. Taken in the year 1928, it was fortuitously in one of a series of several exposures (providing 'before' and 'after' shots for comparison) photographed the same night. Both an analysis of the character of the transient's image in the emulsion and the fact of a rounded shape (compared with the elliptical star images made by motion of the camera) are consistent with a brief duration relative to the 45-minute exposure. Since this discovery, two more archived optical transient effects have been found in other network burst source fields [Schaefer et al., 1984]. These also have decades of separation between the time of the optical flashes, in the early to mid-1900's, and the gamma ray bursts, in 1978 and 1979. Since the celestial positions of the optical transients are known to several seconds of arc, it has been possible to examine those fields to present-day optical limits. The results are that only extremely faint source candidates of inconclusive and possibly time-varying character can be marginally inferred [Pedersen et al., 1983; Schaefer, Seitzer, and Bradt, 1983]. Of course, source proper motion over the decades between optical and gamma ray transients may be a problem in the utility of the archived positions.

Optical transient observations obtained in real time, however, might make possible comparative studies that could be useful for 'instantaneous' gamma ray burst source positions. At present, two new kinds of ground-level instruments are being built in the hopes of exploiting this effect. One will survey and map the sky in optical transient effects [Ricker et al., 1983]. Another will use directional information from the first to reorient a telescope mirror, thus determining a transient's position as it occurs with the maximum precision obtainable, viewing it within its calibrating, neighboring stellar field [Teegarden et al., 1983]. Both these are being installed at Kitt Peak Observatory for operation in the near future.

Another unique course of events prompted by network gamma ray burst observations centers on the March 5, 1979 gamma ray transient. That occurrence provided a picture that still appears, as it did at the time, to be that of a once-in-a-lifetime event. This burst differed in its properties in such detail that I was convinced that it was not 'another' gamma ray burst, but a separate class of event in itself [Cline, 1980]. That claim, made when high resolution burst

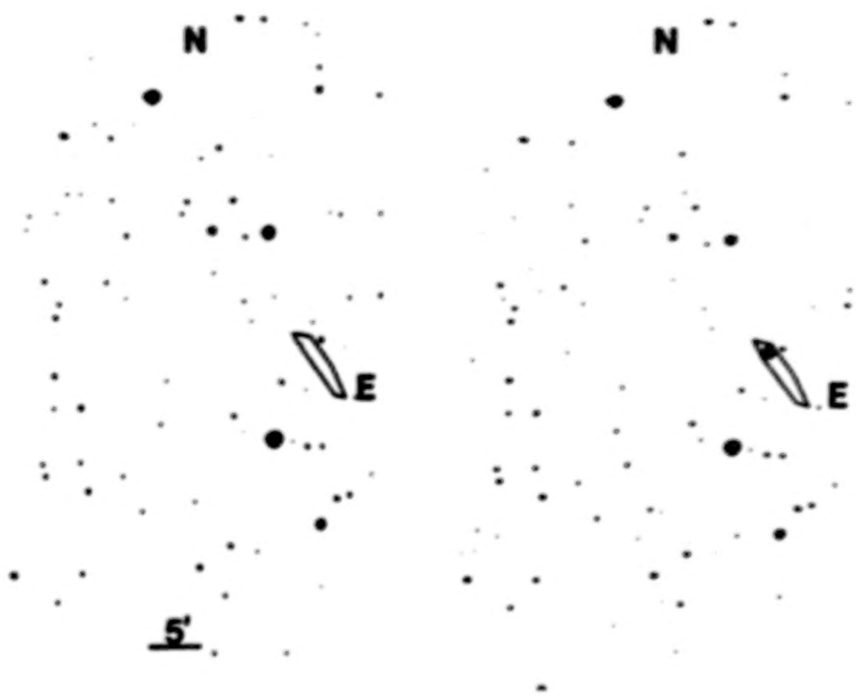


Figure 2. Two 1928 archival plates, including the one with the first optical transient [Schaefer, 1983] found within a gamma ray burst source field [Cline et al., 1984], which is also plotted.

instrumentation had been in operation for three years' time, still appears as a reasonable approximation. No event monitored in the seven years since has come close to duplicating it. The Leningrad group concentrated on the 8-second periodicity as a central feature and found that the continuum spectrum was softer than usual [Mazets et al., 1979a]. It also later discovered the series of small events that appeared to trickle out in sequel fashion [Mazets et al., 1979b]. Instruments on ISEE-3, Helios-2, and Pioneer-Venus Orbiter had the resolution to observe the <0.2 msec risetime of this transient [shown in Figure 3; Cline et al., 1980], which remains particular only to this event of the many hundreds logged. The initial measurement of the source direction [Evans et

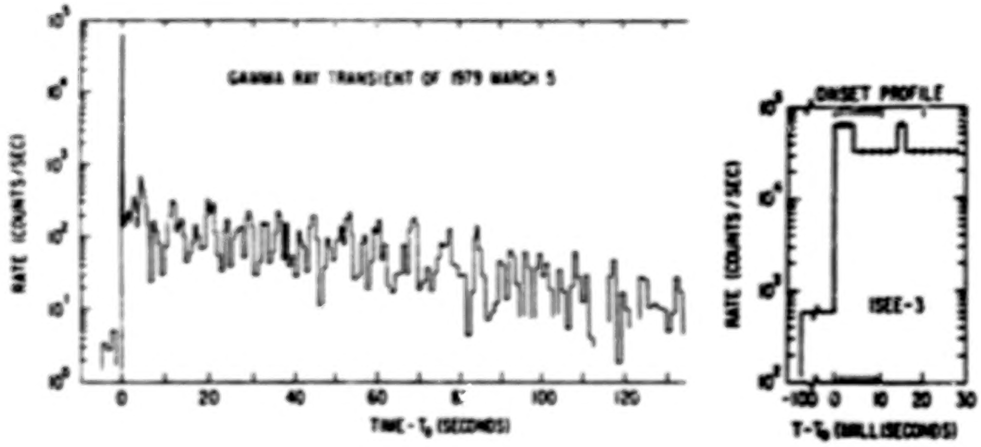


Figure 3. The time history of the March 5, 1979 event [Cline et al., 1980]. On the left is the overall picture, illustrating the intense, 150-msec wide peak and the periodic declining afterglow. On the right is the detail of the unique onset, with its < 0.2 msec rise time constant.

al., 1980] provided the spectacular but controversially interpreted result of a precise, two-arc-minute fit onto the position of N49, a supernova remnant in the Large Magellanic Cloud (LMC) at a distance of 55 kpc, about 5 times the distance from us of the galactic center. A complete analysis using all available measurement capability refined it to a sliver-shaped field only seconds of arc from the center of N49 [Figure 4; Cline et al., 1982]. This remains as the most precise measurement in gamma ray astronomy.

The photon spectrum of the March 5 event, like some others, contained a 400 keV increase [Mazets et al., 1979a]. This experimental feature was barely capable of independent confirmation, using the ISEE-3 high resolution gamma ray spectrometer, only in the case of another event [Teegarden and Cline, 1980]. Its existence remains as a controversial issue, given the lack of confirmation with the SMM spectrometer [Nolan et al., 1983, 1984] in a large number of more recent Venera events. Soon after the March 5 event, its various features, including the line, were fit to models consistent with the N49 distance [Ramaty et al., 1980; Liang, 1981]. One possible explanation for the energy

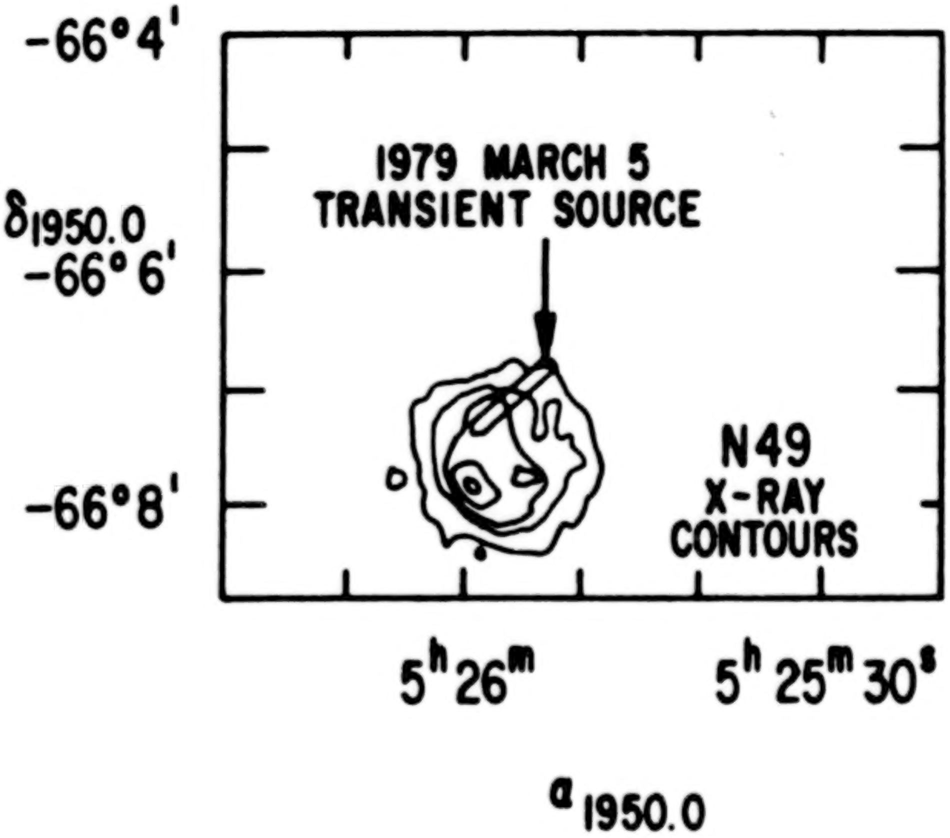


Figure 4. The precise source position of the March 5, 1979 event [Cline et al., 1982], plotted on the contours of the N49 supernova remnant, as measured with the Einstein X-ray telescope [Helfand and Long, 1979].

mechanism used the gravitational storage mode of a neutron star [Ramaty, Lingenfelter, and Bussard, 1981]. This source controversy also persists, with a very recent contention that the distance of N49 may be outside that physically possible by a factor of about 5 unless some gamma ray beaming exists [Liang, 1986]. Such a requirement does not seem a strong constraint, given the rarity of the sole detection.

5. PHENOMENOLOGY

The intent of this note is not to review the field of gamma ray transients, quite the contrary. My purposes in what remains—acknowledging the encouragement of Frank McDonald in the creation of the network—are to tie these examples of the early contributions that the network made to astrophysics together with certain very recent developments (that it also made possible) and to provide a new viewpoint regarding the observations. This recent view [Cline, 1986] appears to be compatible with all the facts and to provide the possible resolution of certain current inconsistencies. It also continues to favor an N49 source for the March 5, 1979 event, the identification, as outlined above, that I have always supported.

The subject of gamma ray bursts has been reviewed in detail in three conferences and workshops in recent years, with published proceedings edited by Lingenfelter, Hudson, and Worrall [1982], Woosley [1983], and Liang and Petrosian [1984]. These reviews virtually exhausted the material then available and are highly recommended. In spite of all the attention to the details of the experimental results, however, very little that is definitive has been produced by theoretical burst studies. This is not entirely the 'fault' of the theoreticians, since (with the exception of the March 5 event) there is no identified candidate source object to provide a source distance, nor is there a source pattern anisotropy to calibrate a scale for the source distances within the galactic disk.

There is another shortcoming inherent in interpretations of continuum burst spectra. I base it on a combination of misfortunes: (1) gamma ray spectra are 'obliging', which means that the observed pulse height distributions cannot be unambiguously converted into energy spectra; (2) burst time histories, from early measurements to the present, are seen to fluctuate dramatically, perhaps beyond the limits of instrumental resolution—and energy spectra are of necessity measured with considerably coarser time-resolution than are time histories—so it is clear that their inferences may be in considerable error [e.g., Norris et al., 1986]; and (3) even minimizing these limitations, observed burst spectra generally happen to be quite amenable to a wide variety of presumably specific fits. Thus, a 'fault' of the model makers may instead be excessive zeal, permitting overconfident interpretations of the experimental details. In

fact, with so little that can be truly pinned down about gamma ray bursts, the range of source ideas seems to have stopped converging but instead now provides some *deja vu* with its variety. One recent cosmological origin model combines the latitude that modern physics and superstrings can give to the imagination with gravitational focusing [Paczynski, 1986].

One gamma ray burst observational puzzle has centered for years on the inconsistency of an observed cutoff in the size spectrum with the fact of an isotropic source distribution, as one sample illustrates in Figure 5. (A plot of the integral number of events 'N(S)' seen with magnitude greater than size 'S' would obey a power law of index -1.5 if the sources are randomly distributed throughout an indefinitely extended three-dimensional region of space; it would taper to an index nearer -1.0 for a population of events coming from a two-dimensional source volume like the galactic disk, in which case an anisotropy could be observed.) A great deal of attention has been devoted to resolving this problem. Approaches range from the adoption of a galactic halo source region [Jennings, 1985], which would be clearly consistent both with isotropy and with a size spectrum cutoff, to the selection of a redefinition of 'size' (using peak, rather than total, intensity) that can be adjusted so as to provide a spectrum with no cutoff problem [Higdon and Lingenfelter, 1986]. Also, it has been popular to simply attribute instrumental inadequacies and miscalibrations as responsible for any observed cutoffs so as to dismiss the problem. The latter permits the view that burst sources are very near by, even compared with the thickness of the galactic disk, a view that always had the intrinsic appeal of minimizing energy considerations. The earliest models required a nearby source from photon density considerations [Schmidt, 1978], although their unconfirmed spectral 1 MeV cutoff is another question.

I have been concerned for some time, however, that the size cutoff problem is more involved than is given credit, with its generally unnoticed connections to two other issues. First, the occurrence of groups of typically small events had been observed on at least two occasions [Mazets et al., 1979b, 1981]. Each presents a cluster in time from 'repeater' sources, i.e., each series has its own mutually consistent source directions. One concern was therefore that if the instruments in use are able to observe small events, identifying them in isolated groups or patterns, and if these are also gamma ray bursts, then those same detectors could not be exhibiting the instrumental insensitivity that appears

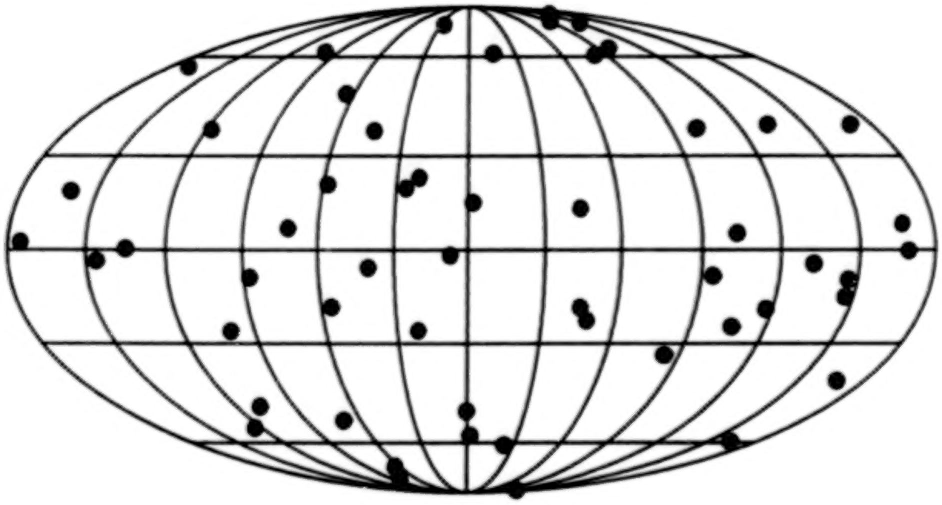


Figure 5. One source distribution pattern of gamma ray bursts [Atteia et al., 1986], seen to be consistent with isotropy. Different instruments produce their individual catalogs; this one was recently compiled with observations from the interplanetary network.

to exist regarding the creation of a cutoff in the size spectrum of the more intense events. The small events should also be observed in greater numbers, in random directions and at random times like the larger bursts, quite clearly not occurring only in specific directions and in groups. One possible resolution to this seemingly trivial inconsistency is discussed below.

The second concern was that the March 5, 1979 event and its properties may be misunderstood, i.e., underestimated in its relevance, in two ways:

(1) The event surely has too precise a positional agreement with the supernova remnant N49 to have that possible identification ignored by writing it off as 'accident'. In spite of the great mathematical unlikelihood of its chance coincidence, a variety of workers in this field have, since its discovery, preferred to work on that event as an unusually bright gamma ray burst considering it to be from a source

possibly closer than most [e.g., Helfand and Long, 1979]. I have felt that this outlook would needlessly waste a singular opportunity to investigate what may be a far more instructive lesson in physics.

(2) The March 5 event was too anomalous in its properties, with its < 200 microsecond onset time constant and its clear periodicity, to be defined as 'another' gamma ray burst. Both of these properties remain as anomalous as ever, with the accumulation of hundreds of additional events for comparison [e.g., Hurley, et al., 1987]. The opposing view is that an economy of assumptions argues against considering the event to be in a distinct class. However, to identify that event as a typical gamma ray burst and interpret its periodicity as confirming a neutron star origin concept for gamma ray bursts always seemed to me surely distasteful if the phenomenon of its periodicity sets it apart from all other gamma ray bursts. That approach has also seemed to me particularly unaesthetic if its source identification with a supernova remnant (an object necessarily associated with a neutron star) is simultaneously dismissed.

Thus, the issue of the gamma ray burst size spectrum was not separable either from the problems of understanding the distance of the most interesting event or from the issue of the number of event classes.

Two other distinctive features of the March 5, 1979 event pertain to the scenario to be suggested. First, as alluded to earlier, it appears to have a considerably softer spectrum than most gamma ray bursts. A large proportion of bursts intense enough to permit accurate differential spectra are characterized in the 150 keV region [Cline and Desai, 1975] with recently found extensions to many tens of MeV [Matz et al., 1985]. The March 5 event, like its associated sequel series and like the other spring 1979 series, is characterized instead in the 30 keV region.

Second, the fact that this intense event can be associated with several other events is both unique and relevant. All hard or 'classical' bursts appear to be isolated in source direction. Low-intensity events followed (and perhaps other similar events may have preceded) the March 5 event. These events were found from 1979 to at least the early 1980's with independently determined

source directions that overlap onto a common field of one or two square degrees [Mazets et al., 1981] implying a common source. That field includes the arc-minute source field of the March 5 event, implying in turn that both it and this series originated from the same source. The positional evidence for that identification is a factor of about 3600 weaker than the positional connection of the March 5 event to N49 (!) but no analogous reasons exist to contest it. The small events have similar maximum intensities but their time histories vary considerably. One flat 'square wave' of 3.5-second duration has no trace of the compound 8-second periodicity so clear in the March 5 event, providing an incidental piece to the puzzle.

6. CONCLUSIONS

A recent discovery is that a third series has been found. This one has a source direction in the galactic bulge at several degrees from the galactic center (close enough to support the assumption that its origin is likely to be at that distance) and consistent with the source direction of an event of four years' earlier observation [Laros et al., 1986]. Some of the generally small and brief events were found buried in 1983 Prognoz-9 data [Hurley, et al., 1987], and were confirmed as well as augmented with a greater profusion of single-spacecraft candidate events in ISEE-3 data [Laros, et al., 1987]; some were also confirmed in SMM data [Kouveliotou, et al., 1987]. The spectrum of the January event is somewhat similar to the spectra of the other two repeater series. All three have spectral characters in the 30 keV region, well above that of X-ray bursts and equally well below the several hundred keV character of the hard gamma ray bursts. As mentioned above, this further distinguishes them from the hard bursts, none of which have yet been found to repeat. The parameter of time history also provides at least a statistically distinguishing feature: the hard bursts can have temporal durations over a very wide range, from the fractional seconds to at least one minute, as well as varying from simple to complex in temporal structure, whereas the soft events are generally brief and simple.

Thus, a new classification of events (occasionally suggested over the years to account for earlier indications of distinctions based on one or another of these three parameters) now appears to be more evident than before. It seems

that this consideration also resolves several features of the size spectrum issue discussed in previous pages. The repeating, soft bursts and the hard, or 'classical' bursts differ considerably in their source and emission properties. Thus, detector effects clearly could provide for a relative sensitivity to one class and a relative insensitivity to the other, producing the 'cutoff' in the classical gamma ray burst size spectrum but having an entirely differing bias for the repeater populations—yet unknown, since their size spectra are entirely uninvestigated. Further, all the small-event repeating series appear to fit this class of events, with characters intermediate in energy at around 30 keV and with basically single-spiked time histories. The March 5 event process itself must relate to the production of that class. Since these intermediate-energy and hard classes of gamma ray transients have so little in common, the fact that the periodicity of the March 5 event may imply its neutron star origin does not necessarily reflect on the origins of 'classical' gamma ray bursts—although it is certainly most likely that all kinds of transients from X-ray bursts to gamma ray bursts do have neutron star origins. The emission processes responsible for these event classes, however, may be surely as distinct as they appear to be.

What I suggest, in concluding, is the possibility of a fourth characteristic that may distinguish these two event classes, namely, source direction pattern. Gamma ray bursts of the hard, nonrepeating, and common variety may have an isotropic source pattern, but the source directions of the three series of soft repeating events can be interpreted to show a glimmer of an emerging pattern. Based on the 'statistics of three counts', of course, it is nevertheless interesting to note that the three repeaters have sources consistent with the galactic plane, the galactic bulge near the center, and N49 in the nearby (by galactic dimensions) LMC.

All this defends the N49 identification with a plausibility picture, based on the occurrence of repeating event origins in high density (galactic or LMC) regions. It is consistent with resolving the several inconsistencies in gamma ray burst phenomonology outlined earlier. As illustrated in Figure 6, this contention makes for a comparison of the familiar source pattern seen in visible starlight, in X-ray binaries [Wood et al., 1984], and in the infrared. Also, the intensities of the galactic bulge events may be assumed to be somewhat greater than those of the sequels from the March 5 source, since they have

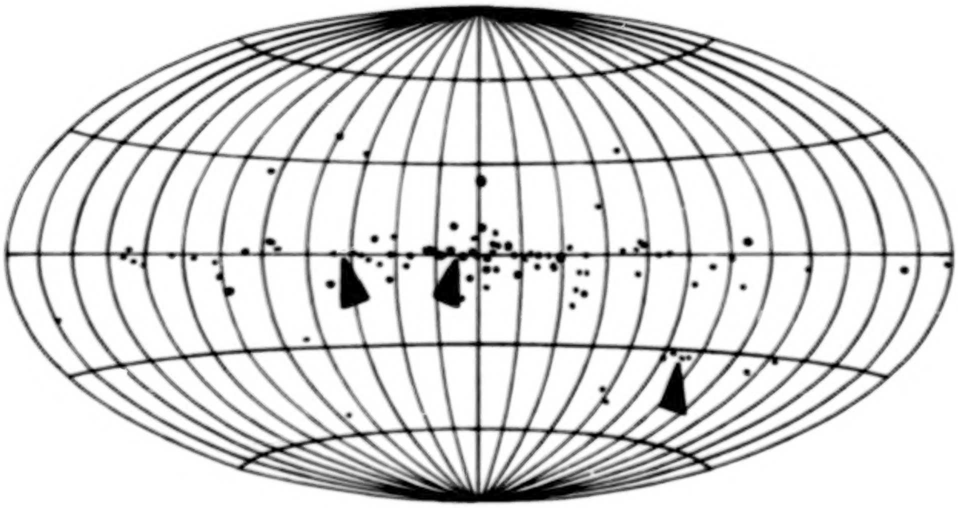


Figure 6. The sources of the three known intermediate-energy, repeating gamma ray transient series. It is too soon to have a statistically meaningful pattern; the three locations available are, like the X-ray binaries, consistent with high density regions in the disk and LMC.

been observed with less sensitive instrumentation. (That is, the only detector capable of detecting the March 5 sequels was not in use for the recent bulge event discovery, whereas some of the instruments observing the recent series had also been up during those years.) That inference is consistent with the fact of the galactic center to LMC distance ratio. The March 5 event itself is, as before, the exception. Perhaps further analyses will provide new enlightenment. In the meantime, the viewpoint suggested may provide a clue towards the understanding of the nature of the singular March 5 event and of gamma ray transients in general.

7. EPILOGUE

Questions form the natural termination of a science essay, rather than a list of accomplishments. The questions-to-answers ratio in the situation regarding gamma ray bursts is larger than that which is typical for a subdiscipline

now over 12 years old. 'Enigmatic' is still one of the adjectives often used and 'puzzle' a frequent noun. Given the continued nonavailability of reliable long duration balloons, it is a fact that all variations of gamma ray transients can be investigated only with spacecraft, although the optical transient connection may someday permit sea level studies to be made. The Gamma Ray Observatory will contain the only 'next-generation' instrument planned to be put in orbit for some time, although several instruments similar to those used in the past decade will continue to see service on nondomestic programs. No plans exist for a spacecraft high resolution gamma ray burst spectrometer that might, for example, separate red-shifted annihilation lines from those expected from a 'grasar', i.e., a gamma ray annihilation line laser [Ramaty, McKinley, and Jones, 1982]. Such are the unanswered fantasies that studies of gamma ray transients can promote. Valuable and hopefully definitive information should surely be forthcoming from GRO. Perhaps the continued scrutiny of existing data will produce additional surprises. Frank McDonald has said, "Are gamma ray bursts a transient phenomenon?" Was the exploration of space, as we knew it in the decades past, with its opportunities for highly individualized creativity, flowering within the massive and seemingly impersonal team projects, a transient phenomenon? One thing is clear: the heyday of space science was both Camelot, as those in Greenbelt happily knew it, and wild frontier, as those who launched their own science payloads in balloon gondolas, rockets, and satellites are privileged to remember. As such, like the physics era of 'string and sealing-wax', it cannot be repeated.

REFERENCES

Atteia, J. L. et al., 1986, *Astrophys. J. Suppl.*, in press.

Barat, C. et al., 1984a, *Astrophys. J.*, **280**, L50.

Barat, C. et al., 1984b, *Astrophys. J.*, **286**, L5.

Cline, T. L., 1980, *Comments Astrophys.*, **9**, 13.

Cline, T. L., 1986, COSPAR Proc., Toulouse, France, in press.

Cline, T. L., and Desai, U. D., 1975, *Astrophys. J.*, **196**, L43.

Cline, T. L. et al., 1973, *Astrophys. J.*, **185**, L1.

Cline, T. L. et al., 1976, *Nature*, **266**, 749.

Cline, T. L. et al., 1979, *Astrophys. J.*, **232**, L1.

Cline, T. L. et al., 1980, *Astrophys. J.*, **237**, L1.

Cline, T. L. et al., 1981, *Astrophys. J.*, **246**, L133.

Cline, T. L. et al., 1982, *Astrophys. J.*, **255**, L45.

Cline, T. L. et al., 1984, *Astrophys. J.*, **286**, L15.

Colgate, S. A., 1968, *Canadian J. Phys.*, **46**, S476.

Evans, W. D. et al., 1980, *Astrophys. J.*, **237**, L7.

Helfand, D. J., and Long, K. S., 1979, *Nature*, **282**, 292.

Higdon, J. C., and Lingenfelter, R. E., 1986, *Astrophys. J.*, in press.

Hurley et al., 1987, in press.

Jennings, M., 1985, *Astrophys. J.*, **295**, 51.

Klebesadel, R. W., Strong, I. B., and Olson, R. A., 1973, *Astrophys. J.*, **182**, L85.

Kouveliotou et al., 1987, in press.

Laros, J. G. et al., 1981, *Astrophys. J.*, **245**, L63.

Laros, J. G. et al., 1986, *Nature*, **332**, 152.

Laros et al., 1987, in press.

Liang, E. P. T., 1981, *Nature*, **292**, 319.

Liang, E. P. T., 1986, *Astrophys. J.*, in press.

Liang, E. P. T., and Petrosian, U., eds., 1984, *Gamma Ray Bursts*, AIP Conf. Proc. **141**.

Lingenfelter, R. E., Hudson, H. S., and Worrall, D. M., eds., 1982, *Gamma Ray Transients and Related Astrophysical Phenomena*, AIP Conf. Proc. **77**.

Matz, S. M. et al., 1985, *Astrophys. J.*, **288**, L37.

Mazets, E. P. et al., 1979a, *Nature*, **282**, 279.

Mazets, E. P. et al., 1979b, *Soviet Astron. Letters*, **5**, 343.

Mazets, E. P. et al., 1981, *Nature*, **290**, 379.

Meegan, C. C., Fishman, G. J., and Wilson, R. B., 1985, *Astrophys. J.*, **291**, 479.

Morrison, P., 1985, *Nuova Cimento*, **7**, 858.

Nolan et al., 1983, AIP Conf. Proc. **101**, eds. M. L. Burns, A. K. Harding and R. Ramaty, 59.

Nolan et al., 1984, *Nature*, **311**, 360.

Norris, J. P. et al., 1986, *Astrophys. J.*, **301**, 213.

Paczynski, B., 1986, *Astrophys. J.*, **308**, L43.

Pedersen, H. et al., 1983, *Astrophys. J.*, **270**, L43.

Ramaty, R., Lingenfelter, R. E., and Bussard, R. W., 1981, *Astrophys. Space Sci.*, **75**, 193.

Ramaty, R. et al., 1980, *Nature*, **287**, 122.

Ramaty, R., McKinley, J. M., and Jones, F. C., 1982, *Astrophys. J.*, **256**, 238.

Ricker, G. R. et al., 1983, AIP Conf. Proc. **115**, ed. S. Woosley, p. 669.

Schaefer, B. E., 1983, *Nature*, **294**, 722.

Schaefer, B. E. et al., 1984, *Astrophys. J.*, **286**, L5.

Schaefer, B. E., Seitzer, P., and Bradt, H. U., 1983, *Astrophys. J.*, **270**, L49.

Schmidt, W. K. H., 1978, *Nature*, **271**, 525.

Teegarden, B. J., and Cline, T. L., 1980, *Astrophys. J.*, **236**, L67.

Teegarden, B. J. et al., 1983, AIP Conf. Proc. **115**, ed. S. Woosley, p. 687.

Wheaton, W. A. et al., 1973, *Astrophys. J.*, **185**, L57.

Wood, K. et al., 1984, *Astrophys. J. Suppl.*, **56**, 507.

Woosley, S., ed., 1983, *High Energy Transients in Astrophysics*, AIP Conf. Proc. **183**.

N87

24261

UNCLAS

14

DISCRETE X-RAY SOURCES AND THE X-RAY BACKGROUND

Riccardo Giacconi
Space Telescope Science Institute
Homewood Campus
Baltimore, Maryland 21218

ABSTRACT

Since the discovery, more than twenty years ago, of a highly uniform X-ray background (XRB) in the 2-10 keV range, its nature has not yet been fully explained. It appears clear from the results of "Einstein" medium and deep surveys that at least 50 percent of the XRB is due to individual extragalactic sources when their contribution is integrated to $Z = 3$. This includes contribution from Quasi Stellar Objects (QSO's), Active Galactic Nuclei (AGN's), galaxies, and clusters of galaxies. The average spectrum of each of the individual contributing sources is softer than that of the observed XRB (power law index $\alpha \approx -0.4$ from 3 to 10 keV). Therefore, the remaining contribution must have a rather hard spectrum of $\alpha \approx 0.0$ - 0.2. It is unlikely that this spectrum can be produced by diffuse processes. Therefore, the remainder of the XRB must be due to individual sources with the appropriate spectrum. This requires either that the spectrum of the already identified sources changes at early epochs or a new class of objects. Advanced X-ray Astrophysics Facility (AXAF) observations will extend survey sensitivity to limiting fluxes of order of 3×10^{-16} erg cm^{-2} s^{-1}, some 50 times fainter than any previous survey. They will have sufficient sensitivity and angular resolution to permit identification and study of these objects.

1. INTRODUCTION

The existence of an isotropic X-ray background with intensity of 1.7 photons $cm^{-2}\ s^{-1}\ sr^{-1}$ (in the 2 to 8A° range), was established in the same flight of June 18, 1962, which discovered the first X-ray star, Sco X-1, Giacconi et al. [1962], Figure 1.

Due to the high degree of isotropy, it became evident, even after the first few observations, that the X-ray background had to be mainly extragalactic in origin. If extragalactic and due to a uniform distribution of sources, a substantial fraction of it (> 20%) had to originate at large distances (Z > 1), and therefore would carry information of cosmological interests. The very first theoretical paper after the discovery of X-ray sources, by Hoyle [1963], pointed out that the hot steady-state cosmological model of Hoyle and Gold predicted

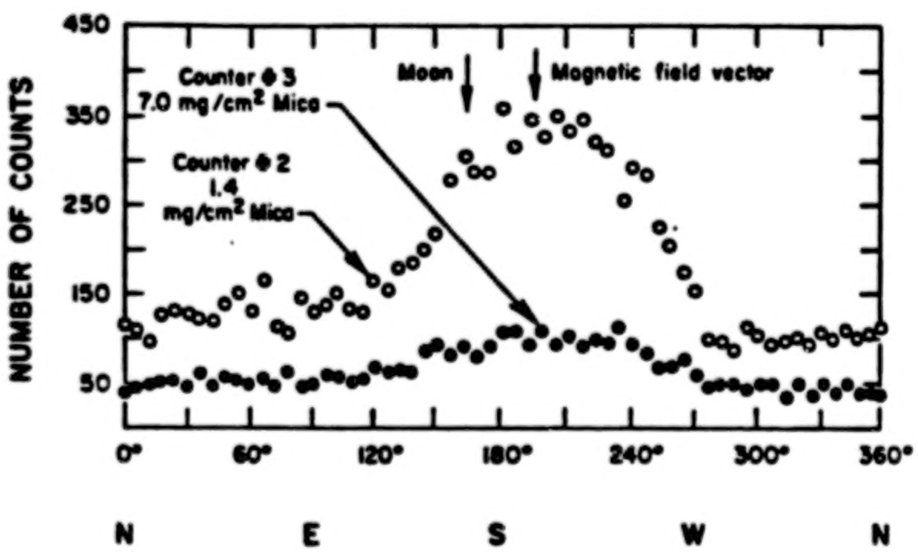


Figure 1. Azimuthal distributions of recorded counts from Geiger counters flown during June 1962. Giacconi, R. and Gursky, H., eds., X-ray Astronomy, D. Reidel Publishing Company, 1974.

a greater flux of background X-rays than was observed. This single observation marked the beginning of the demise of the steady-state theory. More important was that it demonstrated the potential of X-ray observations in studying the early universe.

In the twenty years since then we have learned a great deal about the spectrum and angular distribution of the X-ray background. It is generally accepted that about half of it is the result of the superimposed emission from many distant individual extragalactic sources: normal galaxy, low luminosity active galaxies, QSO's, and clusters of galaxies all contribute to the observed emission. As to the question whether the X-ray background can all be explained in this manner, the answer is "no". If the remainder of the background is due to individual sources they must differ in their emission spectrum and/or evolution from any of the known classes. Perhaps protogalaxies or early stages of QSO's may provide this emission. While there is no evidence for a truly diffuse component, a contribution as large as 10-20% cannot be excluded.

2. EARLY OBSERVATIONS AND THEORY

Opportunity for indepth study of the nature and origin of the background came from the "Uhuru" mission, Figure 2. It is clear from the figure that, in X-rays, the background radiation dominates the night sky, a result quite different qualitatively from what one obtains in the visible range of wavelength. From the study of Uhuru data, it could be determined that fluctuations of the background were less than about 3% on angular scales of 10 degrees. This was strong confirmation of the extragalactic origin of the background. It was shown by Schwartz and Gursky [1974] that under reasonable assumptions on the density and composition of intergalactic gas, optical depth of unity would not be reached in the X-ray regime, until $Z \approx 7$, even in the hypothesis of a closed universe. This means that X-rays generated at very early epochs can reach us unimpeded by absorption or scattering effects. If we now ask from what range of distances may the radiation actually be coming, we find that this depends on assumptions about the emissivity function. For a uniform emissivity (in co-moving coordinates) throughout the universe, one finds that some 20% of the background must come from $Z > 1$; if we assume any kind

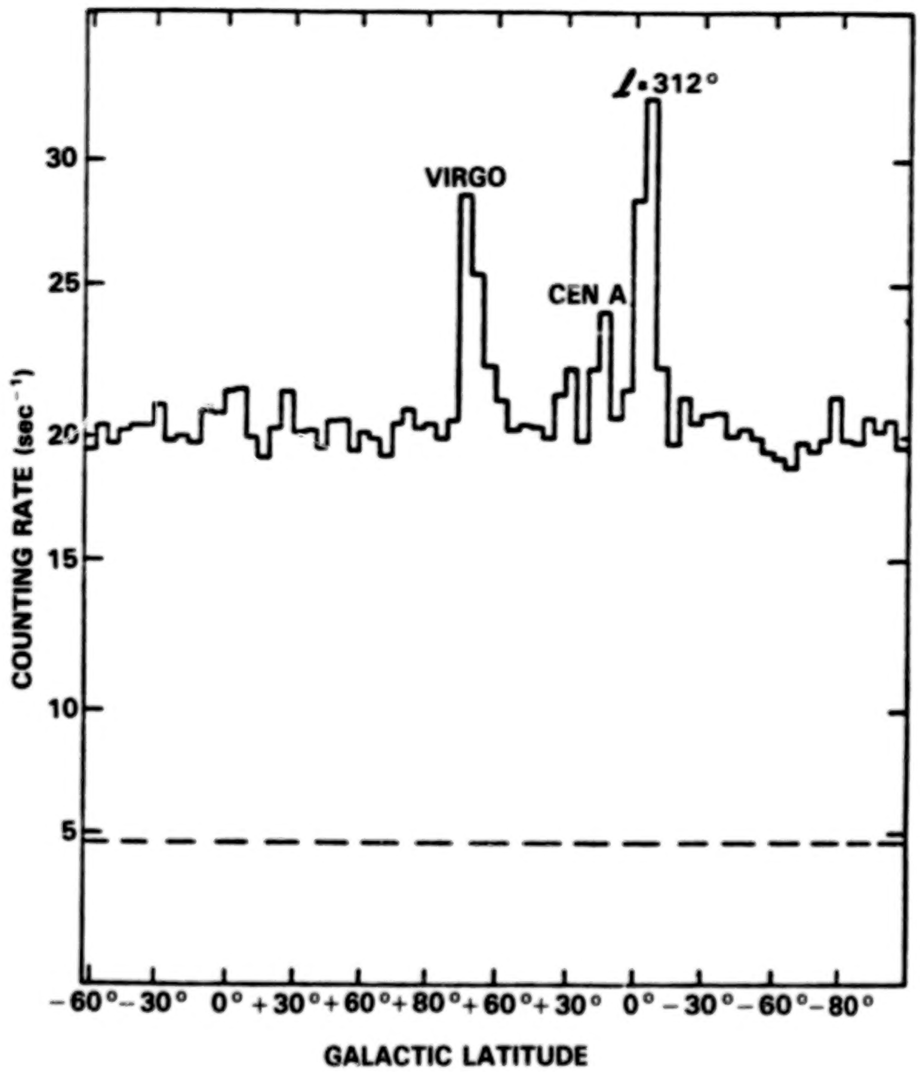


Figure 2. Uhuru counting rates along a great circle passing through high galactic latitudes. Except for discrete sources, the telescope records a general radiation background which is not related to the galaxy. The small contribution of non-X-ray background is indicated by the dashed line. Giacconi, R. and Gursky, H., eds., X-ray Astronomy, D. Reidel Publishing Company, 1974.

of evolution (the simplest being an additional factor of $(1+Z)^3$ as would be expected if the emissivity depends on ρ^2) then we find that most of the background must originate at $Z > 1$, Figure 3.

As to the spectrum of the observed radiation, an early compilation by D. Schwartz of results from a number of groups is shown in Figure 4. Schwartz concluded that the spectrum could be best fit by two power laws of the form

$$I(E) = \begin{cases} 8.5\ E^{-0.40} & 1 \leqslant E \leqslant 21 \text{ keV} \\ \\ 167\ E^{-1.38} & E \geqslant 21 \text{ keV} \end{cases}$$

with I(E) in keV/keV cm^2 s sr or by an exponential fit of the form $I(E) = 4.1 \exp\text{-}(E/35)$ as would be expected from thermal emission by a gas with effective temperature of 4×10^8 K. He noted that an exponential spectrum provided a poor fit both at low energies (< 1 keV) and at very high energies (> 300 keV).

A number of theories were proposed to explain the background radiation origin in the early 1960's and 1970's. The first proposal, already mentioned, was by Hoyle [1963] of X-ray production from a hot intergalactic medium, whose existence had been predicted by Gold and Hoyle [1959] as a consequence of matter formation in their steady-state cosmological model. It turned out that this model would predict a flux one hundred times greater than observed! Felten and Morrison [1966] suggested that inverse Compton scattering of cosmic ray electrons on the 3 K radiation may reproduce the power law spectral shape. If the intergalactic electrons could be assumed to have the same spectrum as cosmic electrons in our galaxy (spectral index $\alpha \approx 2.6$), then a spectrum of the type $E^{-1.5}$ could be obtained. The main difficulty with this model is the lack of detailed knowledge on cosmic electron fluxes at early epochs. Then current estimates yielded a derived X-ray flux one or more orders of magnitude less than that observed.

Cowsik and Kobetich [1972] proposed that the spectrum could be interpreted as a thermal bremsstrahlung from a gas at a temperature of several hundred

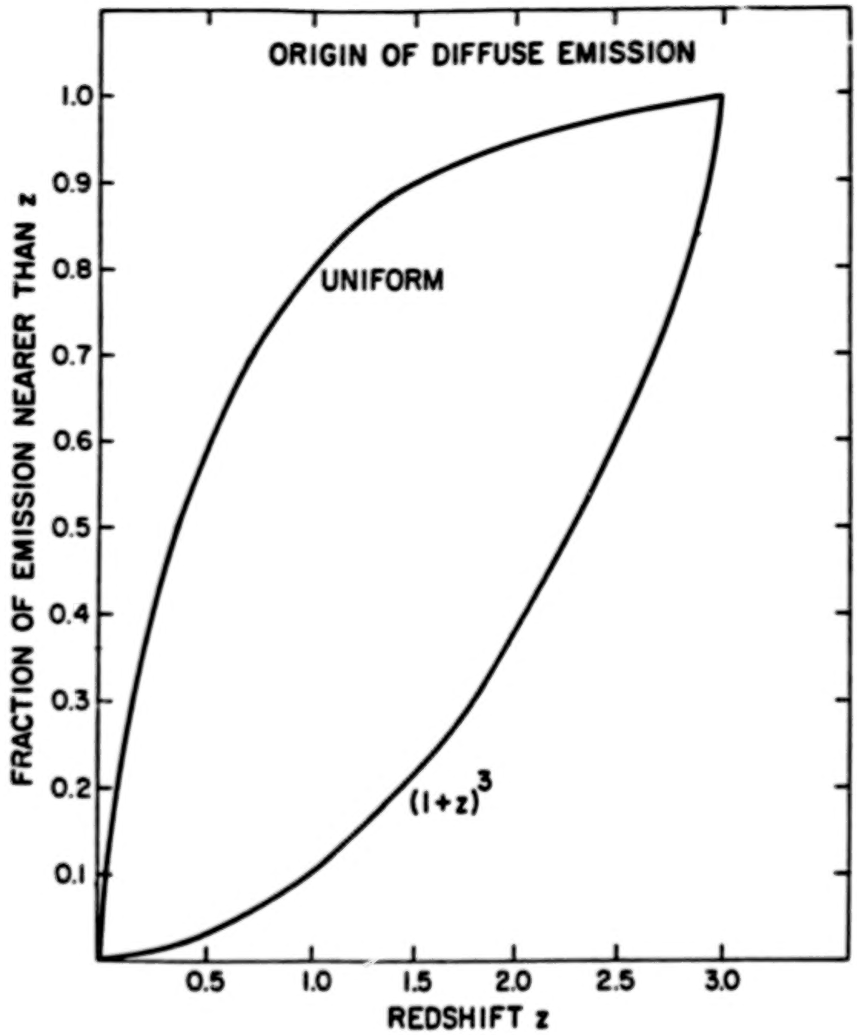


Figure 3. Fraction of the X-ray background which is produced nearer than the indicated redshift. Two models are shown: uniform emissivity (in comoving coordinates; upper left curve) and emissivity which evolves by an additional factor $(1 + z)^3$ (lower right curve). In both cases we assume $q_o = 0.5$, the X-ray energy spectral index is -0.5, and the emission is cut off at $z = 3$. Schwartz, D. A., 1978, Proc. IAU/COSPAR Symposium on X-ray Astronomy, Innsbruck, Austria.

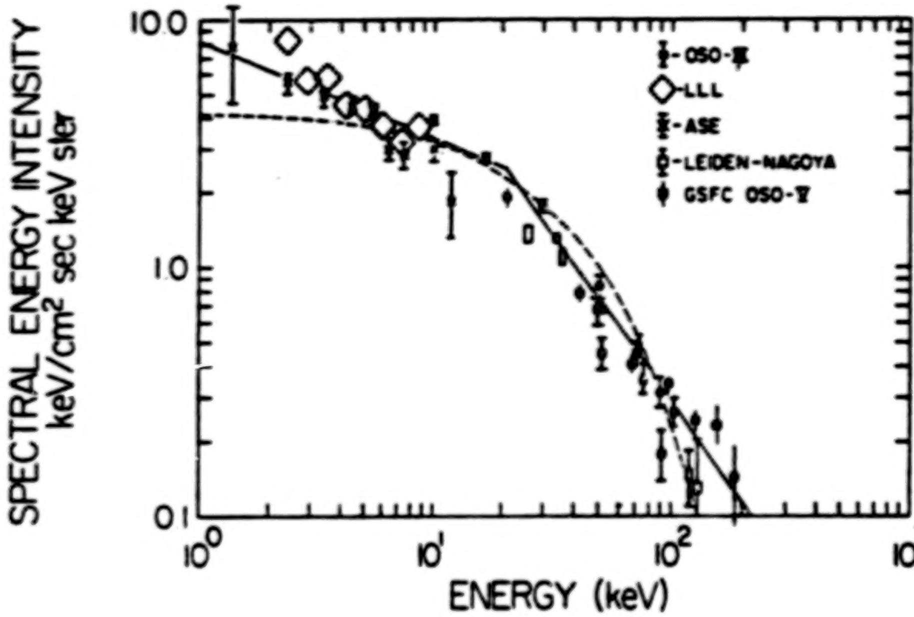


Figure 4. Measurements of the diffuse X-ray energy spectrum in (keV keV^{-1} cm^{-2}sr^{-1}s^{-1}). Data were selected from published results according to the following criteria: (1) The experiment either spanned a wide range of geomagnetic conditions, or (2) included means of directly detecting and reducing effects of charged particles in the detector. Results reported only in the form of parameters for an assumed spectral shape are excluded. The data presented here have sufficient precision to reject the hypothesis that a single power law shape fits all the data between a few keV and a hundred keV. OSO 3: SD 70b, SD 73a. LLL:PT 71. ASE:GP 69. Leiden-Nagoya: BJ 70. OSO5: DB 73. Schwartz, D. A., 1978, Proc. IAU/COSPAR Symposium on X-ray Astronomy, Innsbruck, Austria.

million degrees. They computed a detailed fit of the model to the existing spectral data, Figure 5. They obtained a temperature of 3.3×10^8 K, $\int N_e N_p dl \sim 10^{17}$ cm^{-5}, $\rho \approx 3 \times 10^{-6}$ H atoms cm^{-3} and $\rho \approx \rho_{crit}$ if $H_o = 55$ Km s^{-1} Mpc^{-1}.

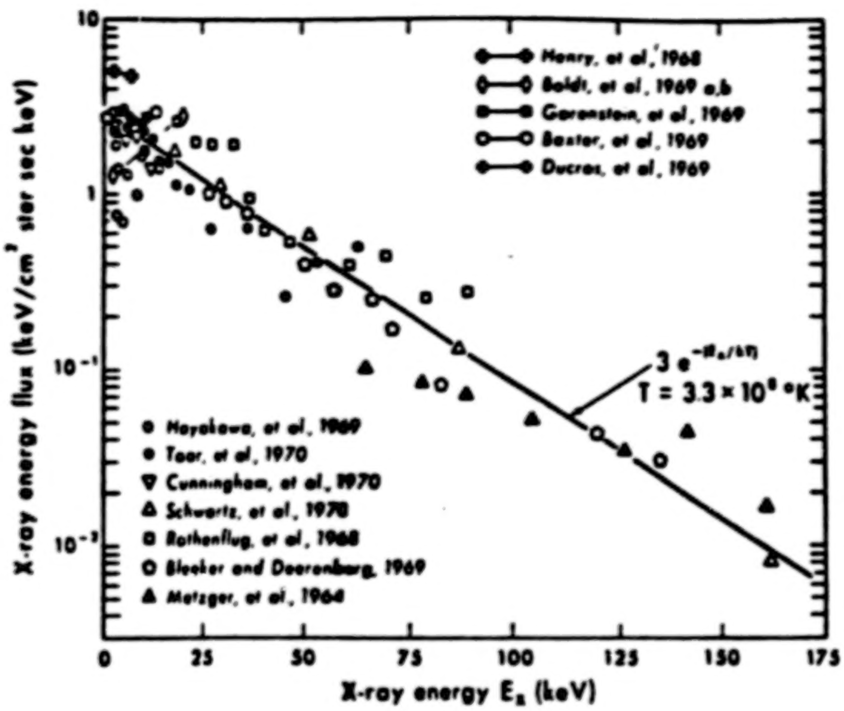


Figure 5. The difference between the observed energy flux and the calculated flux in the energy interval 2 keV E_x 200 keV is plotted as a function of X-ray energy. The line represents the thermal-bremsstrahlung emission for a hydrogen plasma at 3.3×10^8 K. The line-of-sight integral $\int N_e N_p\, dl$ for this emission is 1.3×10^{17} cm^{-5}. If one assumes no clumping and $\int dl \approx 2c/3H_o \approx 10^{28}$ cm, one gets $N_e \approx N_p \approx 3 \times 10^{-6}$ cm^{-3}. Such a density is adequate to close the Universe if $H_o \approx 55$ km s^{-1} Mpc^{-1}. Cowsik and Kobetich, 1972, Astrophys. J., 177, 585.

Field and Perrenod [1977] gave the most complete and critical treatment of this model. They start from the fit of Cowsik and Kobetich. They compute the energy content of the gas which turns out to be larger than any other form of energy in the universe except that in the 3 K or in the rest mass. They pose the question of how the heating of the gas could occur. They suggest that

an early population of QSO's with numbers evolving as $(1+Z)^6$ might be the source. They assume that heating occurred at $Z=3$ and that adiabatic cooling has taken place since, Figure 6. Then they can provide a description of → and under the assumption that $\rho \approx \rho_o (1+Z)^3$, they find that the integrals from $Z=0$ to $Z=3$ would yield a fit with $T = 4.4 \times 10^8$ K and $\Omega = 0.46\ C^{-1/2}$ (where C is the clumping factor $C = \langle n^2 \rangle / \langle n \rangle^2$). They remain skeptical of this explanation because of the very large energy requirements, the large value of Ω (in baryonic matter) and the existence of diffuse HI clouds in intergalactic space which would evaporate in the presence of the postulated hot intergalactic medium, [Mattewson, Cleary, and Murray, 1975; Cowie and McKee, 1977].

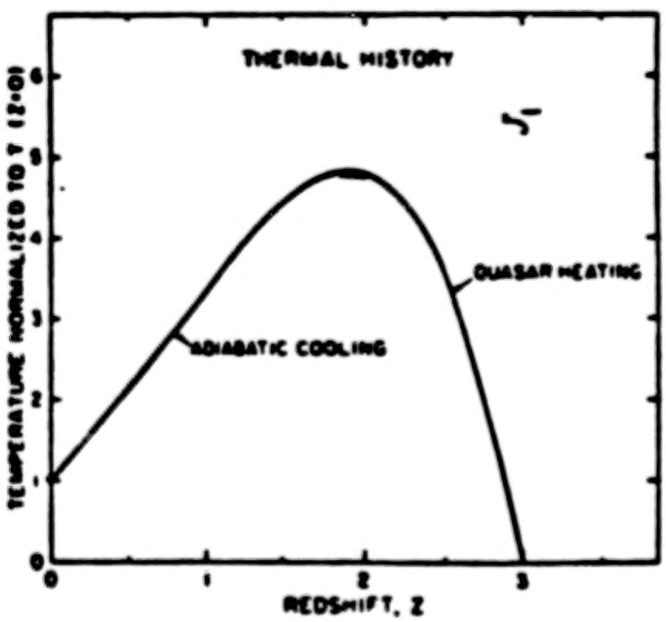


Figure 6. The variation of the temperature of intergalactic gas with redshift. The temperature scale is in units of its present value, T_0, which is determined to be 4.4×10^6 K from the observations. Field and Perrenod, 1977, Astrophys. J., 215, 717.

3. EARLY LOG N-LOG S DETERMINATIONS AND THE CONTRIBUTION OF INDIVIDUAL SOURCES

Most of the theories discussed up to now attempted to explain the background as a truly diffuse emission from the Intergalactic Medium (IGM). However, starting in the early 1970's, a body of experimental evidence became available regarding the X-ray emission from extragalactic sources and also regarding the number count of extragalactic sources as a function of detected flux. These findings shifted the question of the nature of the X-ray background to a more fundamental issue, namely, whether the background was truly due to diffuse processes or to the unresolved contributions of individual sources. The salient facts were the findings that the Number-Intensity distribution for high galactic latitude, presumably extragalactic, sources ($b \geq 20°$) in the Uhuru catalog followed a power law distribution -1.5, while for galactic sources ($b \leq 20°$), it followed a -0.4 distribution, Figure 7, [Matilsky, 1973]. This was interpreted as the result of a luminosity function independent of distance, in which case the extragalactic sources much nearer than $Z = 0.1$ will dominate the counts and follow the classical $N(>S) = KS^{-3/2}$ law.

It therefore became possible to speculate that the additional contribution of fainter sources (below the Uhuru limit of 1.7×10 ergs cm^{-2} s^{-1}) might give rise to the remainder of the background. Setti and Woltjer [1970] had, in fact, pointed out as early as 1970, when only 3C273 had been observed as an X-ray source, that the integrated contribution of all QSO's could easily explain the entire X-ray background.

In a pre-"Einstein" review, Schwartz [1978] analyzed the contribution to the background from known extragalactic sources. He took into account the known X-ray luminosity function for Seyferts, clusters of galaxies, normal galaxies, and QSO's. He concluded that each class contributed several percent to the background without luminosity or density evolution. No class of discrete X-ray source could constitute the entire background without evolutionary effects, but less pronounced evolution for QSO's could account for all of the X-ray background. New data on individual sources at fainter limits would be required to settle the question.

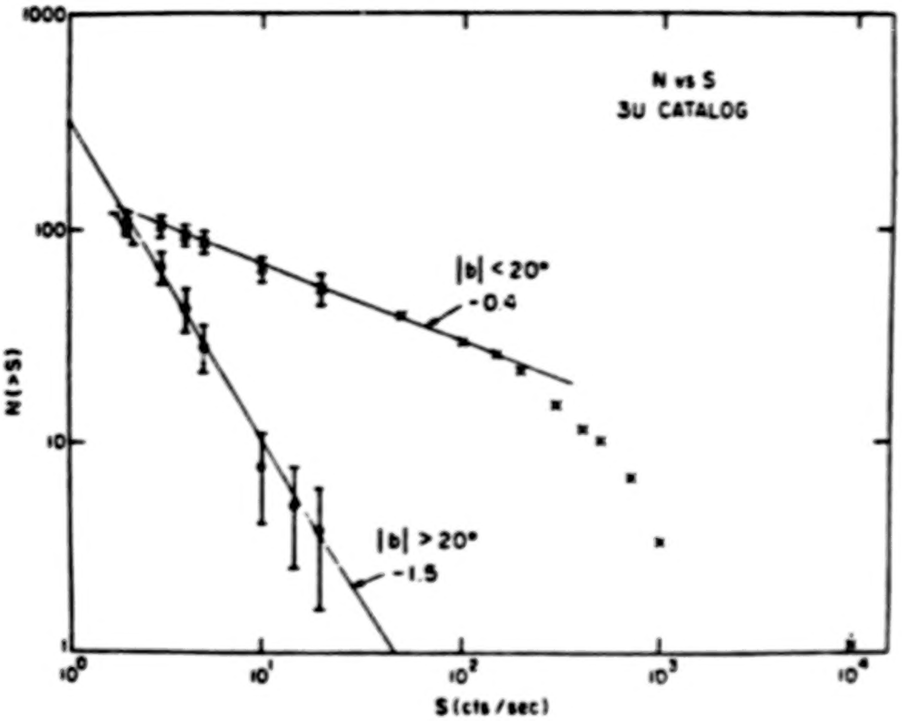


Figure 7. Number-Intensity distribution for sources in the UHURU catalog from Matilsky et al. 1973. Giacconi and Gursky, 1974, X-ray Astronomy.

4. THE HEAO-1 FINDINGS ON SPECTRAL SHAPE

Diffused emission theories of the X-ray background and thermal bremsstrahlung models in particular received new impetus after the findings of the HEAO-1 (A-2 experiment), which obtained X-ray background spectra in the 2 to 40 keV region with high statistical accuracy, Figure 8.

Marshall et al. [1980] showed an excellent fit to the Field and Perrenod model with parameters of $Z_{cutoff} = 3$, $T_o = 26$ keV and $\Omega = .36C^{1/2}$. They repeated the Schwartz arguments against the explanation of the background as due

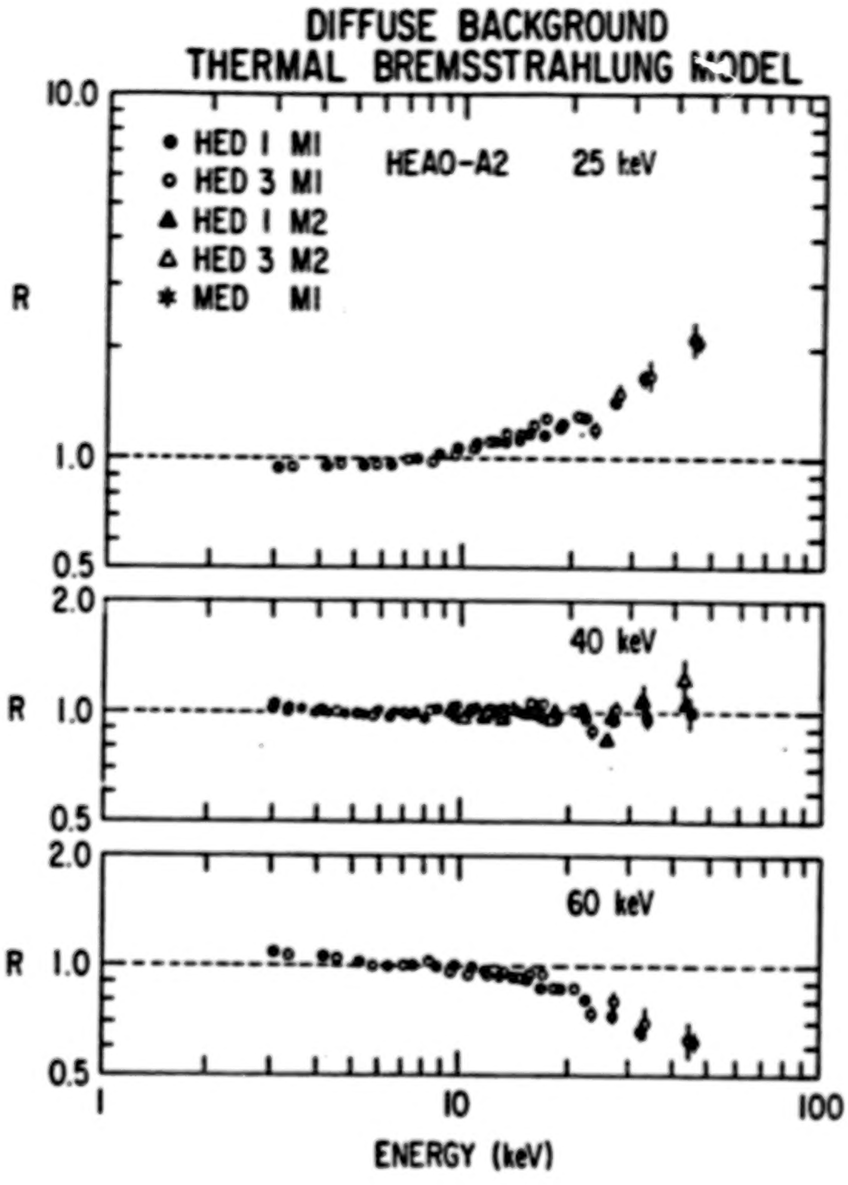


Figure 8. Marshall, F. et al., 1980, Astrophys. J., 235, 4, and Boldt, E., 1981, Comments on Astrophysics, 9, 97.

to discrete sources without evolution. They further questioned whether an evolving population (QSO's) would have the right spectrum to mimic thermal bremsstrahlung emission. They concluded that the measured spectrum strongly suggested that a substantial fraction of the background is due to a hot intergalactic gas with near closure density.

Several arguments have been raised against this interpretation. These are: (1) the argument of Field and Perrenod based on energetics; (2) the high density of the IGM in baryonic matter which is in contradiction with the value of $\Omega \approx .1\,\Omega_{crit}$ derived from deuterium abundances; (3) the existence of HI intergalactic clouds, which would evaporate in the presence of this postulated hot intergalactic medium; (4) the argument first proposed by Cavaliere and Fusco-Femiano [1976] and more recently discussed by Forman, Jones, and Tucker [1984] based on the existence of clusters; and (5) the realization that such good fits only hold over the restricted energy range 2 to 40 keV. At energies > 100 keV, the spectrum steepens significantly requiring, in the thermal bremsstrahlung interpretation, a higher temperature component of $\leq$ 95 keV, Schwartz [1978].

5. DIRECT MEASUREMENTS OF DISCRETE SOURCE CONTRIBUTIONS FROM "EINSTEIN" (HEAO-2)

The Einstein Observatory was used to study the contribution of discrete extragalactic sources to the background in a number of ways.

There are basically three methods to determine the contribution of a class of discrete sources to the background from the "Einstein" data:

(1) Direct measurement of Log N-Log S and identification of the extragalactic nature of the sources in the 0.5 to 3 keV energy range.

(2) Measurement of the X-ray luminosity function for a specific class of objects and knowledge of its evolution.

(3) Measurement of optical properties (luminosity function and evolution) of a specific class of objects and knowledge of their Lx/Lopt ratio.

All these methods have been employed by analysis of the Einstein data and have resulted in mutually consistent estimates of the contribution of sources to the background.

Method 1 was implemented through the deep surveys [Giacconi et al., 1979] which resulted in extending the Log N-Log S measurements by about three orders of magnitude from the Uhuru limit of 10^{11} erg cm^{-2} s^{-1}, 1-3 keV range, to 1.3×10^{-14} erg cm^{-2} s^{-1}. The medium Einstein survey [Maccacaro et al., 1982] filled the gap between these extremes. From integration of the Log N-Log S relation a background contribution of $\geq 35\%$ could be directly computed [Murray, 1981], Figure 9.

Method 2 was implemented by use of the medium survey data to obtain a local luminosity function for AGN's and the evolution parameter c in the assumption of a pure luminosity evolution [Maccacaro et al., 1983]. (The assumption of a density evolution could be shown to be inconsistent with the data.) Maccacaro estimated a contribution to the background of the evolving AGN's of 78 $^{+35}_{-17}$ %. A lower limit to this contribution consistent with the range of parameters if 42%; the upper bounds exceed 100%.

Method 3 was implemented by using Einstein data on QSO's X-ray luminosity, determination of the ratio L_x/L_o and its dependence on L_o [Zamorani, 1982; Avni and Tananbaum, 1982; Maccacaro et al., 1983], and use of the optical evolutionary models to compute the contribution of QSO's to the background (BKG). Marshall reports a value of 75 $^{+66}_{-32}$ % [Marshall, 1983].

It is worth noting that Methods 2 and 3 are quite uncertain because they are to a degree dependent on a number of parameters describing the luminosity function, the limits of integration L_x min and L_x max, the evolutionary laws of the specific class of objects (for instance QSO's), as well as on the parameters of the assumed world models. It should also be noted that a substantial fraction of the sources that would give rise to the background have not been directly observed.

Furthermore, as pointed out by De Zotti et al. [1982] and Boldt and Leiter [1984], the sources which dominate the X-ray background cannot have spectra similar to the power laws measured locally for AGN's. It follows that the

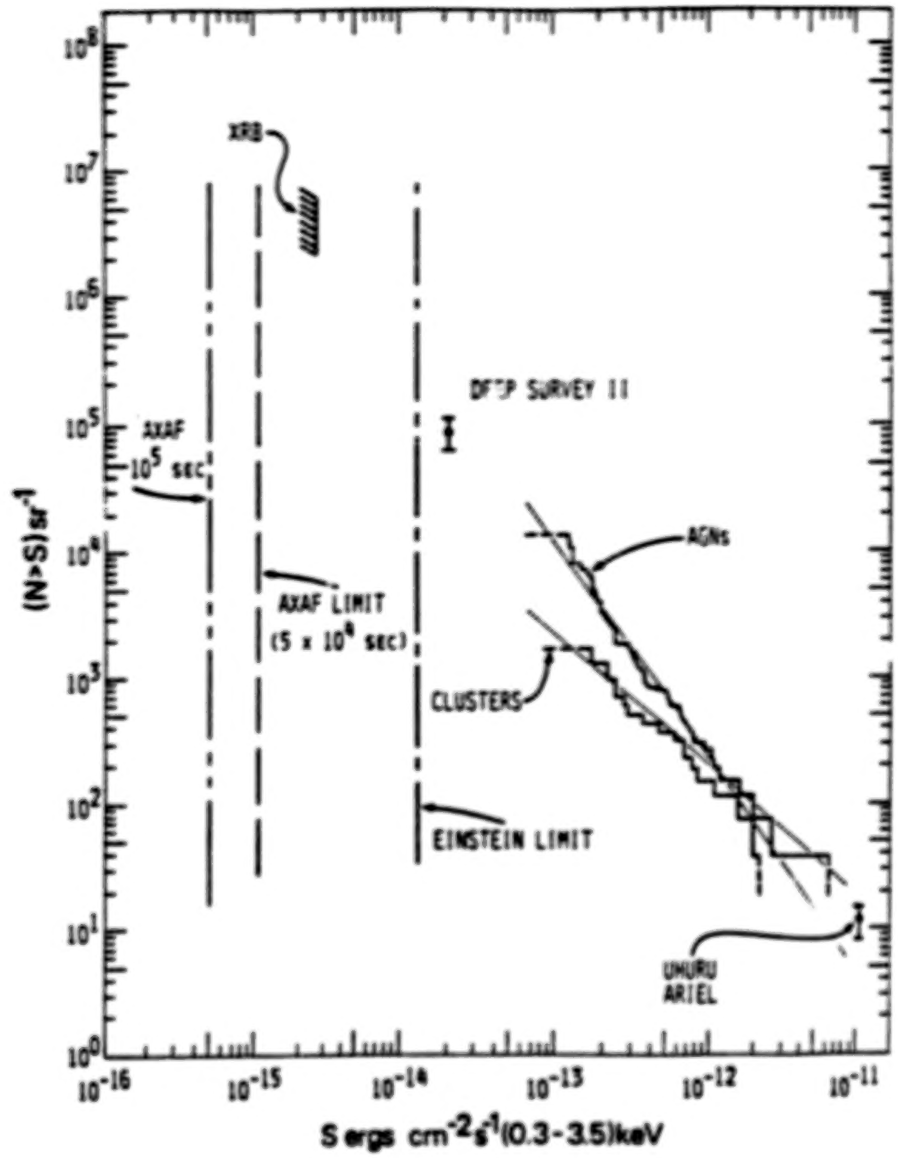


Figure 9. Shows Log N versus Log S as a function of S. The results from previous surveys are shown as well as the estimated limits of the AXAF survey here proposed.

objects must be either QSO's with characteristics different from that of the nearer QSO's or a different class of objects altogether (protogalaxies?). The fact that the spectrum must change renders the extrapolations necessary in Methods 2 and 3 even more uncertain.

The only certain assessment of the contribution of discrete sources to the X-ray background comes, therefore, from the extension of Log N-Log S to the Einstein limit. As to the nature of the objects, the current program of Deep Survey source identification yields mainly QSO's with $0.5 \leqslant Z \leqslant 2$ as optical counterparts of the extragalactic component.

6. CURRENT PROBLEMS AND POTENTIAL FOR FUTURE CLARIFICATION

The current situation can be summarized as follows:

(a) It is certain that a very substantial fraction of the X-ray background is due to the contribution of individual sources. The most complete current evaluation of the contribution from the different classes of objects is that of Schmidt and Green [1986]. The contributions to be expected as a function of increasing sensitivity of the surveys are shown in Figure 10.

(b) There is no direct evidence for the existence of a true diffuse component. Current estimates set an upper limit of 10-20%. These limits come from the current estimates of baryonic matter content from the existence of diffuse HI clouds in intergalactic space and from the survival of clusters; and may be somewhat relaxed by the recent arguments of Fabian on the effect of considering mildly relativistic effects on the gas emissivity [Fabian, 1984]. In order to invoke such effects, Fabian must postulate emission of gas with temperature ~ 200 keV at $Z \sim 5$. (See comment below.) It should be noted that the fact that some 50% of the background is due to known individual sources makes it very unlikely that the remainder of the BKG can be explained as thermal bremsstrahlung emission from a hot thin gas. As noted by many authors [Boldt

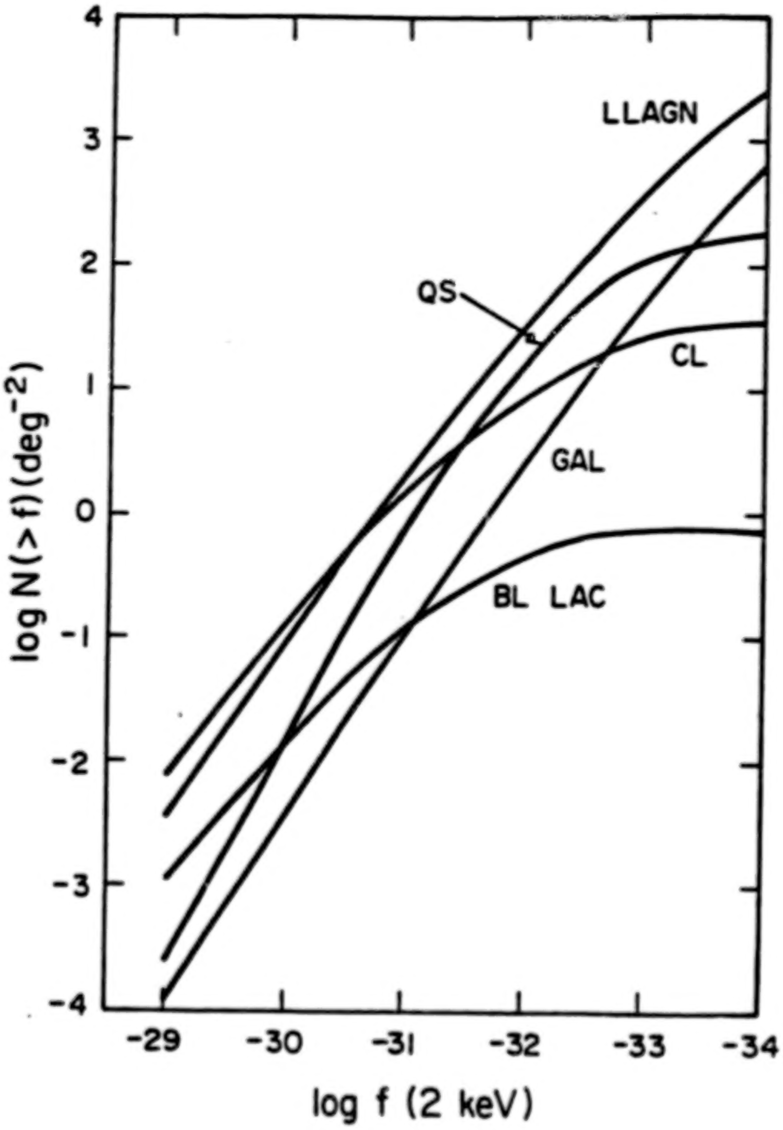


Figure 10. Schmidt, M., and Green, R. F., 1985, "Counts, Evolution and Background Contribution of X-ray Quasars and Other Extragalactic X-ray Sources," preprint.

and Leiter, 1984; Schmidt and Green, 1986], the spectrum of the remainder of the BKG, after subtraction of the known individual source contribution which is softer than the background, becomes very hard. Power laws of index $\pi < 0.2$ are required. In the thermal bremsstrahlung model, this would require $\geqslant 200$ keV at the source and, in order to obtain an effective K $\rightarrow$ at the current epoch of ~ 40 keV, one would require that the emission occur at Z $\gtrsim 6$ [Boldt and Leiter, 1984]. The existence of a heating mechanism for such a high temperature gas at such an early epoch is unknown and, perhaps, not entirely plausible.

(c) Therefore, it appears most likely that the remainder of the X-ray BKG originates in compact extragalactic sources of some kind capable of giving rise to the required hard spectral component. This requires a different emission spectrum from members of a known population (QSO's?) at an early epoch and/or evolutionary effects or a new class of objects (protogalaxies?). The study of this subject, therefore, offers a potential for exciting new discoveries. The most promising approach appears to be a refinement of the direct measurement of individual source contribution at ever smaller limiting fluxes, with high resolution instruments capable of directly imaging the background.

AXAF offers a very substantial opportunity to resolve this problem once and for all. By repeating the deep surveys with increased sensitivity it can extend Log N-Log S to fluxes more than one order of magnitude weaker than detected in Einstein, from 2×10^{14} to 10^{-15} erg cm^{-2} s^{-1} in the 1-10 keV range in 5×10^{4} sec. It can reach 5×10^{-16} erg cm^{-2} s^{-1} in the 1-10 keV range in 10^{5} sec. It can do so over fields of view of the same order as those used in the Einstein HRI deep survey provided that the HRI which ultimately is adopted has the sensitivities recently reported [Fraser and Pearson, 1983] or that a large format charge coupled device (CCD) is selected. Extension of the survey limit to 5×10^{-16} – 10^{-15} erg cm^{-2} s^{-1} (1-10 keV) brings us to a range of fluxes such that we must observe a turnover in the Log N-Log S relationship in order not to exceed the X-ray background. In addition, the AXAF angular resolution of a 0.5 arc-sec will provide the necessary precision to carry out a program of source identifications. It should be noted that this precision is

essential. It was only after the positions for Einstein HRI sources were refined to a few arc-seconds that one and only one optical counterpart would be identified for all X-ray sources. [Some counterparts were found in J plates at magnitudes 20 to 23, Griffith et al., 1983.]

If Lx/Lopt ratios observed in Einstein are maintained for the sources found in these surveys, it follows that when we increase the X-ray sensitivity by factors of 20-40, we will need to observe optical counterparts fainter by three or four magnitudes. We can expect, therefore, that the optical counterparts at the AXAF survey limit will be in the range of m = 24-27. Detection and identification of these objects will require in all probability use of the space telescope (ST) imaging and spectroscopy instruments.

The expected returns from this survey are:

(a) A new determination of Log N-Log S for extragalactic X-ray sources to Smin $\approx 5\times10^{-16}$ erg cm^{-2} s^{-1} (1-10 keV).

(b) A complete X-ray selected sample of 500 to 2500 X-ray sources with fluxes between 5×10^{-16} and 2×10^{-14} erg cm^{-2} s^{-1} (1-10 keV), per square degree surveyed [see for instance estimates derivable from Schmidt and Green, 1986]. This sample will probably consist of QSO's at high redshifts, or clusters or protoclusters and possibly protogalaxies.

Additional work on this survey might include:

(1) Measurement of source position and extent [clusters should have core radius of ~ 20 arc-sec, Bahcall, 1979].

(2) X-ray spectral measurements of individual sources to determine spectral shape for point sources.

(3) X-ray spectral measurements of individual extended sources to determine the presence of emission lines, in particular redshifted Fe Ka emission.

(4) Optical identification and broadband spectroscopy of the candidate objects.

REFERENCES

Avni, Y., and Tananbaum, H., 1982, *Astrophys. J.*, **262**, L17.

Bahcall, N., 1979, in *Scientific Research with the Space Telescope: IAU Colloquium 54*, ed. M. S. Longair and J. W. Warner, NASA CP-2111.

Boldt, E., 1981, *Comments on Astrophysics*, **9**, 97.

Boldt, E., and Leiter, D., 1984, *Astrophys. J.*, **276**, 427.

Cavaliere, A., and Fusco-Femiano, R., 1976, *Astron. & Astrophys.*, **49**, 137.

Cowie, L. L., and McKee, C. F., 1977, *Astrophys. J.*, **211**, 135.

Cowsik, R., and Kobetich, E., 1972, *Astrophys. J.*, **177**, 585.

De Zotti, G. et al., 1982, *Astrophys. J.*, **253**, 47.

Fabian, A. C., 1984, in *X-ray and UV Emission from Active Galactic Nuclei*, ed. W. Brinkmann and J. Truemper (Garching: Max Planck Instituet fuer Extraterrestrische Physik).

Felten, J. E., and Morrison, P., 1966, *Astrophys. J.*, **146**, 686.

Field, G. B., and Perrenod, S. C., 1977, *Astrophys. J.*, **215**, 717.

Forman, W., Jones, C., and Tucker, W., 1984, *Clusters of Galaxies as a Probe of the Intergalactic Medium*, preprint.

Fraser, G. W., and Pearson, J. F., 1984, *Nucl. Instrum. and Methods*, **219**, 199.

Giacconi, R., Gursky, H., Paolini, F., and Rossi, B., 1962, *Phys. Rev. Letters*, **9**, 439.

Giacconi, R., et al., 1979, *Astrophys. J.*, **234**, L1.

Giacconi, R., and Gursky, H., eds., 1974, *X-Ray Astronomy* (Dordrecht: D. Reidel Publishing Co.).

Gold, T., and Hoyle, F., 1959, *Cosmic Rays and Radio Waves as Manifestations of a Hot Universe: IAU Symposium*.

Griffith, R. E., et al., 1983, *Astrophys. J.*, **269**, 375.

Hoyle, F., 1963, *Astrophys. J.*, **137**, 993.

Maccacaro, T. et al., 1982, *Astrophys. J.*, **253**, 504.

Maccacaro, T. et al., 1983, *Astrophys. J.*, **266**, L73.

Marshall, F. E. et al., 1980, *Astrophys. J.*, **235**, 4.

Marshall, H. L. et al., 1984, *Astrophys. J.*, **283**, 50.

Matilsky, T. et al., 1973, *Astrophys. J.*, **181**, 753.

Mattewson, D. W., Cleary, M. N., and Murray, J. D., 1975, *Astrophys. J.*, **195**, L97.

Murray, S., 1981, in *X-ray Astronomy with the Einstein Satellite*, Vol 87, ed. R. Giacconi (Dordrecht: D. Reidel Publishing Co.).

Schmidt, M., and Green, R. F., 1986, "Counts, Evolution and Background Contribution of X-ray Quasars and Other Extragalactic X-ray Sources," preprint.

Schwartz, D. A., 1978, in *Proceedings, Symposium-A on X-ray Astronomy, IAU/COSPAR (Innsbruck, Austria)*.

Schwartz, D. A., and Gursky, H., 1974, in *X-ray Astronomy*, ed. R. Giacconi and H. Gursky (Dordrecht: D. Reidel Publishing Co.).

Setti, G., and Woltjer, L., 1970, *Astrophys. Space Sci.*, **9**, 185.

Zamorani, G., 1982, *Astrophys. J.*, **260**, L31.

N87
24262
UNCLAS

N87-24262

15

THE COSMIC X-RAY BACKGROUND

Elihu A. Boldt
Laboratory for High Energy Astrophysics
Goddard Space Flight Center
Greenbelt, Maryland 20771

1. OVERVIEW

Electromagnetic radiation provides the only channel presently available for obtaining direct information about extragalactic reaches of the universe; additional direct modes of making extragalactic observations (e.g., neutrino astronomy, gravitational radiation) are yet to be exploited. While most studies to date have concentrated on the individually discernible sources of this electromagnetic radiation, those absolute measurements of sky surface brightness providing the totality of the flux in various spectral bands from *all* sources in the universe (i.e., arising from diffuse emission as well as unresolved discrete objects) present us with a most fascinating and challenging constraint. This integrated radiation, known as the "background", gives us a quantitative measure for evaluating how complete our knowledge of the observable universe actually is. Fundamental questions may be posed. Does the total signal observed equal the sum of all the individual sources that we know about from limited samples? Do we have the correct arithmetic (i.e., cosmology) for carrying out this addition? Are there unfamiliar kinds of sources (e.g., with drastically different spectra) that contribute to the background substantially, but have not yet been resolved? What is the motion of our Milky Way galaxy relative to the proper frame of the background? With the advent of comprehensive data from the pivotal HEAO (High Energy Astronomy Observatory) program fostered by Frank McDonald and his associates [see history by Tucker,

1984], the cosmic X-ray background discovered by Giacconi et al. [1962] has now become a leading test case for attacking such questions.

Our present knowledge about the overall spectrum of the isotropic extragalactic background of electromagnetic radiation is summarized in Figure 1. It extends from radio frequencies below the turnover at about 6 MHz (i.e., wavelengths greater than 50 m) to gamma rays above the spectral break at several MeV (i.e., photon wavelengths smaller than 10^{-13} m). The power law observed in the radio band from about 10 MHz to 300 MHz [Clark, Brown, and Alexander, 1970], may be extrapolated to the high energy end of the spectrum [Boldt, 1971, 1974], merging remarkably well with the gamma ray background at a few hundred keV. The background above a few hundred keV may be accounted for in terms of the emission from Seyfert galaxies [Bignami et al., 1979; Boldt, 1981; Rothschild et al., 1983], with no appreciable contribution from sources at high redshifts ($z > 1$) required. In the X-ray band these sources exhibit power law spectra of energy index $\alpha = 0.7$ [Mushotzky, 1982], compatible with the index describing the power law connection of the background between radio frequencies and gamma rays (denoted by the dashed line in Figure 1). The observed spectral break in the gamma ray background at several MeV is taken to be indicative of a characteristic break in the spectra of Seyfert galaxies [Bignami et al., 1979]. The turnover in the radio background spectrum at about 6 MHz [Alexander and Clark, 1974] probably results from the combined effect of free-free absorption in the galactic disk, in the intergalactic medium, and in the sources themselves plus processes such as synchrotron self-absorption and Razin effect [Razin, 1960] in the individual sources. The power law portion of the radio background is well explained as the superposition of radio galaxies and other discrete radio sources [Simon, 1977], mainly associated with the present epoch ($z < 1$). Therefore, from radio to hard X-rays, we appear to have a unified power law background spectrum arising from the integrated contributions of nonthermal sources within the present epoch. This power law, extending over more than 14 decades in photon energy, serves as a baseline for examining additional background components of more limited bandwidth (e.g., "thermal bumps").

The isotropic extragalactic nonthermal radio background referred to above is in fact extracted from an anisotropic galactic background that dominates

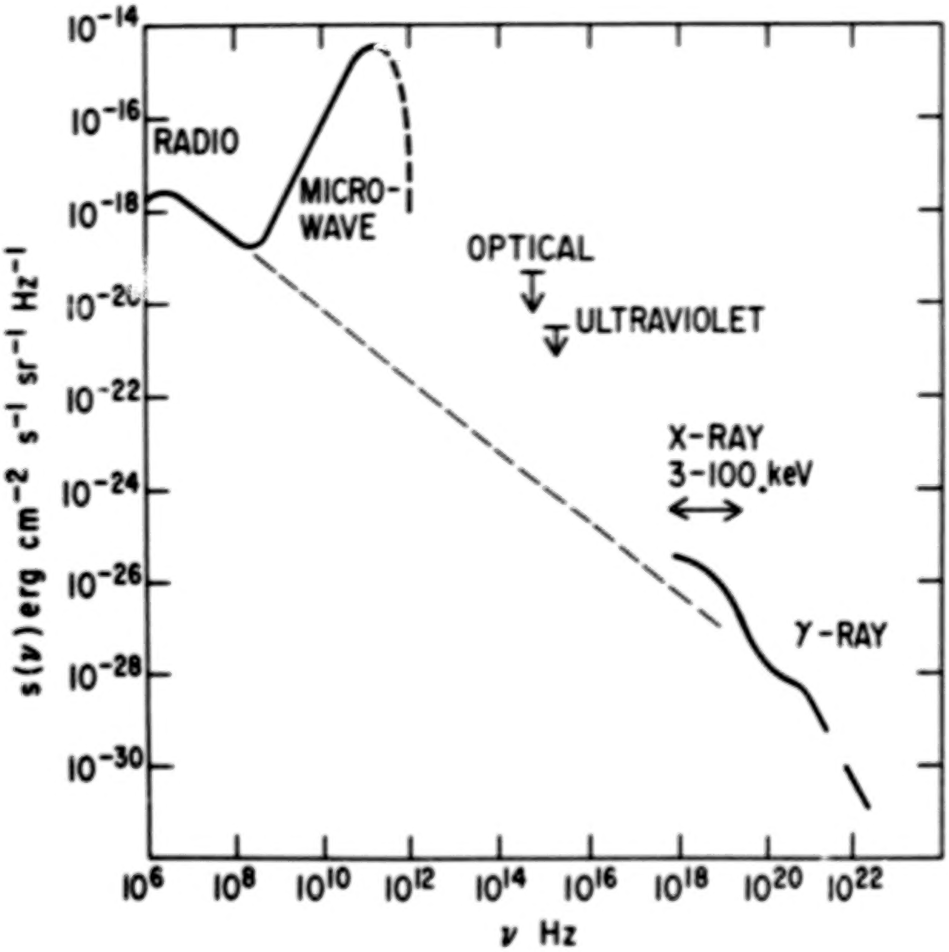


Figure 1. The isotropic sky flux s(ν) versus frequency ν [from Shafer, 1983]. The dashed line extends the power law spectrum at radio frequencies to higher photon energies ($E = h\nu$) assuming $s(\nu) \propto \nu^{-\alpha}$, with $\alpha = 0.7$. References are: Clark, Brown, and Alexander [1970] for radio; Weiss [1980] for microwave; Dube, Wickes, and Wilkinson [1979] for optical; Paresce, McKee, and Bowyer [1980] for UV; Marshall et al. [1980] and Rothschild et al. [1983] for X-rays; and Fichtel and Trombka [1981] for gamma rays.

the total flux. From about 1000 MHz to 10^6 MHz (0.3 mm wavelength), however, the situation changes drastically in that the sky is here dominated by an isotropic radiation field having the spectrum of a blackbody at 2.7 K [see review by Weiss, 1980], generally regarded as the cooled relic of the hot photon gas that dominated the universe after the initial few minutes [however, see Segal, 1983 for an alternate view]. At still higher frequencies (shorter wavelengths) we expect additional thermal components of the extragalactic background corresponding to dust (in the infrared [IR]), starlight (in the optical and ultraviolet [UV]) and possibly from hot gas (in the UV and extreme ultraviolet [XUV]). Upper limits to the extragalactic flux are shown in Figure 1 for the optical [Dube, Wickes, and Wilkinson, 1979; Code and Welch, 1982; Toller, 1983; Weller, 1983] and ultraviolet [Paresce, McKee, and Bowyer, 1980; Joubert et al., 1983]. In general, these upper limits are an order of magnitude higher than the flux anticipated from the superposed contributions of unevolved normal galaxies [Code and Welch, 1982]. Not until X-ray energies above 3 keV do we again encounter an isotropic background component that dominates the sky and where galactic contamination is negligible [for a description of the softer galactic component see McCammon et al., 1983]. This isotropic hard X-ray component extends out to almost 300 keV, appearing as a thermal-type bump well above the power law baseline shown in Figure 1. As pointed out by De Zotti [1982], the precision with which the spectrum of this cosmic X-ray background component has been measured with HEAO-1 [Marshall et al., 1980] exceeds that with which the cosmic microwave background spectrum has been defined. We examine some implications of the spectrum and (an)isotropy of this well-defined X-ray background within the context of individual source spectra, HEAO-2 "Einstein Observatory" source counts [Giacconi, this volume] and the dipole anisotropy of the microwave background. Open questions are emphasized.

2. ROLE OF THE HEAO PROGRAM

Although X-ray astronomy certainly flourished during the initial 15 years of its history, research on the cosmic X-ray background remained comparatively dormant. During this period, experiments concentrated on well isolated bright sources, particularly compact objects of high astrophysical interest, such as

neutron stars in binary stellar systems. However, there were also strong practical reasons for the relatively small attention given to the cosmic X-ray background during the early years of X-ray astronomy. With X-ray telescopes based on the mechanical collimation of photons, rather than focusing X-ray optics, it was not possible to detect any of the faint discrete sources contributing to the unresolved background [for a review of the early technical status see Giacconi, Gursky, and van Speybroeck, 1968]. Furthermore, all detectors used in X-ray astronomy suffer from an annoying background of signals arising from causes other than the X-ray sky; this extraneous background is the basic limiting factor in making the absolute measurements of surface brightness needed for properly defining the cosmic X-ray background. For discrete source measurements, even the "true" diffuse X-ray sky is an undesirable contamination. Hence, prior to the use of focusing X-ray optics, the standard approach was to make the mechanical collimation angle defining the detector aperture as small as practical and to simply measure the increase in signal as the instrument scanned across the discrete source of interest.

It remained for the High Energy Astronomy Observatories (HEAO-1 and HEAO-2) to significantly remedy the relatively poor observational situation that existed before concerning the X-ray background. The all-sky survey carried out over a broad band (from 0.1 keV to over 0.5 MeV) with the HEAO-1 mission (Frank McDonald, project scientist), launched in 1977, involved newly developed experiments especially designed to unambiguously distinguish the X-ray sky background from that due to other causes [Peterson, 1975; Boldt et al., 1979]. The focusing X-ray telescope flown on the HEAO-2 mission, usually known as the Einstein Observatory, brought the power of imaging optics to X-ray astronomy [Giacconi et al., 1979a]. For soft X-rays (< 3 keV), the imaging detectors at the focus of this grazing incidence telescope were used to resolve a substantial portion of the background into discrete faint sources [Giacconi et al., 1979b].

Some of the causes of the extraneous background associated with on-orbit X-ray detectors are charged cosmic rays, ambient electrons, gamma rays (primarily flux as well as a locally generated flux produced by cosmic ray interactions with the spacecraft), and radioactivity induced by passage through the South Atlantic Anomaly [Peterson, 1975]. The strategy for extracting the

diffuse flux of celestial X-rays from the overall background which was employed with the Cosmic X-ray Experiment (A2) on HEAO-1 [Boldt et al., 1979] is summarized as follows:

(1) The most fundamental aspect of a diffuse flux is that it increases linearly with solid angle. In order to apply this definition reliably there were two different rectangular fields of view associated with each detector; one of them was always $3° \times 3°$, the dual one being $3° \times 1.5°$ for some detectors and $3° \times 6°$ for others (where the collimation angle transverse to the scan path was always $3°$).

(2) X-rays are detected via the photoelectric effect, which is a strong function of photon energy (i.e., the cross-section is roughly proportional to E^{-3} and changes at K absorption edges by an order of magnitude). To exploit this pronounced X-ray characteristic there were three classes of gas proportional counters, optimized for low, medium, and high energy X-rays, designated LED (low energy detector), MED, and HED respectively. Propane gas was used for soft X-rays (0.1-3 keV), argon for medium energy X-rays (1.5-20 keV), and xenon for hard X-rays (2-60 keV). [See detailed description by Rothschild et al., 1979.]

(3) Unless there is a bright source in the detector's field of view, most of the raw signals generated by the sensor are due to penetrating charged cosmic ray particles that transverse the sensitive gas volume with a continuous trail of ionization. Since the electrons emerging from a photoelectrically ionized atom are of short range, the secondary ionization resulting from X-ray absorption is very localized. Hence, multi-anode veto logic could be used to efficiently reject cosmic ray signals without rejecting X-ray signals.

(4) Ambient electrons associated with a low earth orbit (such as used for HEAO) are not as penetrating as cosmic rays and are relatively difficult to reject [Holt, 1974]. For a high energy gas proportional chamber, the main xenon-filled detector was separated from the entrance window by an electron sensing veto layer (propane-filled) transparent to hard X-rays. For the low energy detectors (propane-filled) this was not feasible

and deflection magnets were used for removing low energy electrons [effectiveness is described by Garmire, 1979].

(5) Since the scan period of HEAO-1 (0.5 hour) was not small compared with the orbit period (1.5 hours) spatial and temporal effects needed to be separated; this was done by offsetting some of the detectors by an angle of only six degrees along the scan path.

The basic construction of a HEAO-1 (A2) proportional chamber involves an internal multi-wire electrode arrangement that is electrically equivalent to a rectangular stack of independent tubular "wall-less" counters, each with its own anode. X-rays enter via a mechanical collimator on top of the housing (see Figure 2), traverse a thin window and are detected in the multi-counter gas volume of the proportional chamber. The dual collimator is matched to the internal multi-counter structure of the proportional chamber; counters in odd numbered columns are aligned with the $3° \times 3°$ collimator while those in even numbered columns are aligned with the complementary collimator. A cross-sectional photograph of the MED dual collimator is shown in Figure 3 in order to indicate the cellular structure used to implement this scheme.

The HEAO-1 (A2) experiment consisted of six proportional chambers, two LED's for low energies, one MED for midband coverage, and three HED's for hard X-rays. To get some notion of how this experiment covered the sky we can examine the scan path on the celestial sphere for any one of the six detectors; this is shown in Figure 4. Every half-hour this dual collimator detector scanned a complete great circle on the celestial sphere in an angular band of 3° FWHM (full width at half maximum) normal to the HEAO spin axis. At any instant it viewed an angle θ (FWHM) along the scan path via those internal counters associated with the small collimation and an angle ($2 \times \theta$) via the dual portion of the instrument. The spin axis defining the scan plane pointed to the Sun, thereby moving one degree per day along the ecliptic equator. In this way the entire sky was scanned in six months.

The on-orbit performance of the dual collimator scheme is exhibited in Figure 5, where the effectiveness is quite evident. The two histograms displayed give the observed population distribution for count-rate samples sorted according

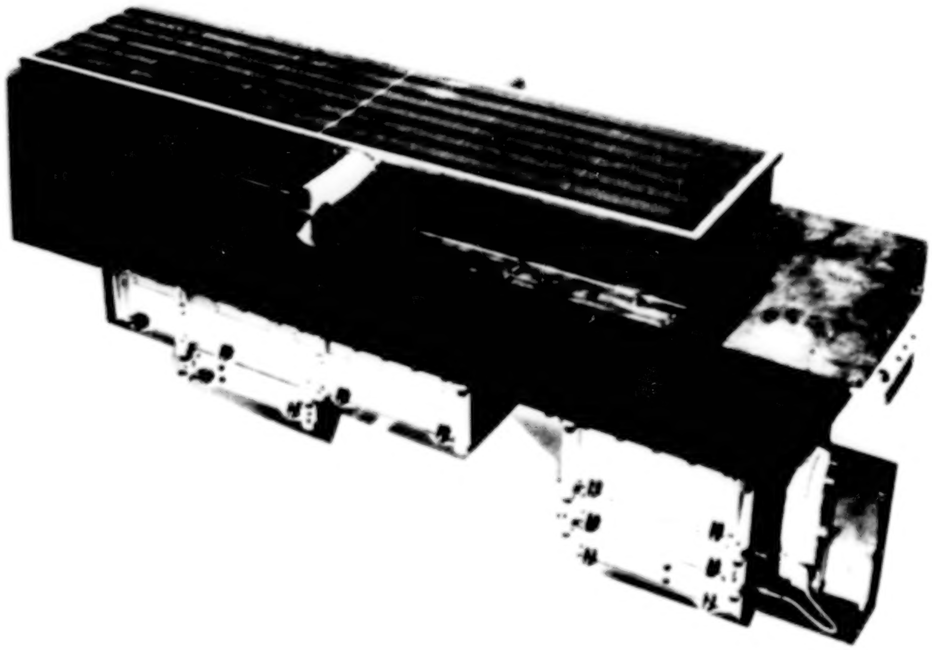


Figure 2. Photograph of a medium energy detector (MED) for the HEAO-1 (A2) experiment; overall length is about one meter. X-rays enter via mechanical collimator on top of housing. Boxes attached to bottom of detector housing are for associated electronics.

to the total number of accepted X-ray events per telemetry major frame (40.96s) recorded with HED-1. They are classified only with regard to field of view (FOV), the top histogram for counts associated with $3° \times 6°$ collimation and the bottom for counts associated with the $3° \times 3°$ collimation. These histograms are based on data obtained over many scan cycles regardless of what was in the field of view, be it Earth albedo or celestial sources. The histogram for each of the fields of view exhibits two clearly separated peaks, the high one attributed to exposures dominated by the sky and the one with lower counts attributed to exposures dominated by Earth albedo. If there were no extraneous sources of background the two histograms would scale as the

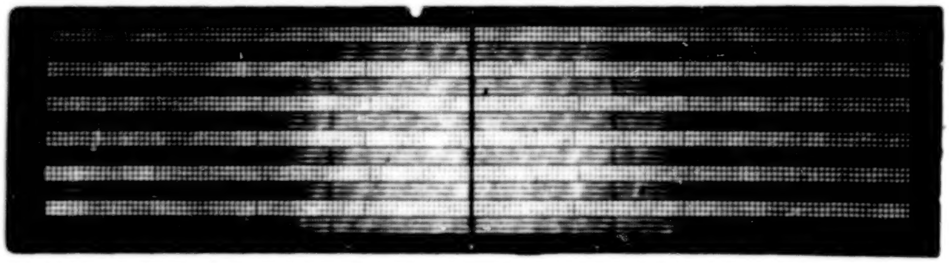


Figure 3. A cross-sectional photograph of the MED collimator showing cellular structure providing 3° × 3° and 3° × 1.5° dual collimation.

ratio of solid angles. In fact, Earth albedo represents a relatively weak source of X-ray surface brightness, even in the hard X-ray band considered here, and most of the signal when Earth fills the field of view arises from a residual background internal to the detector (i.e., due mainly to Compton electrons produced within the chamber by penetrating gamma rays). More extensive data bear out the qualitative indication in Figure 5 that the internal background to be associated with the two fields of view are equal, as expected. Furthermore, a detailed comparison of the internal background derived from the two peaks associated with diffuse sky background has been shown to be the same as that derived from the two peaks associated with Earth background, both in magnitude and spectral shape (i.e., using the relation: (Internal Background) = 2 × (Small FOV) − (Large FOV)). As shown in Figure 5 by a dashed line, the internal background for HED-1 represents an average contamination of only about 14% for the large field of view, even when the full bandwidth (~ 60 keV) of this detector is considered. To exclude effects of internal background completely, however, the spectrum of X-rays to be associated with sky background was obtained by simply subtracting the raw spectrum observed with the small field of view from that obtained with the large field of view for the same detector.

At X-ray energies approaching 100 keV and above, the quantum efficiency obtained with gas counters is just not high enough and "tricks" of the sort

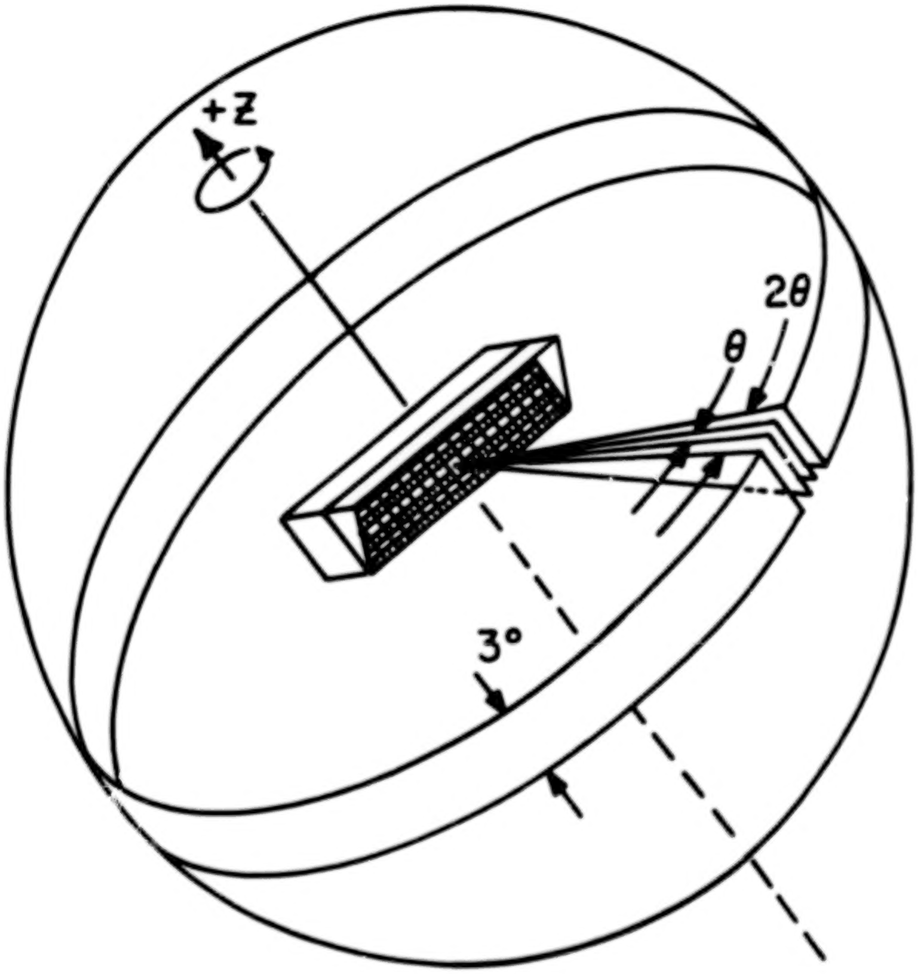


Figure 4. A schematic representation of the HEAO-1 celestial scan as depicted by one of the six (A2) detectors. The scan axis (z) always points to the Sun; as shown, each detector has a sun shade for shielding the collimator from solar radiation. The dual collimation angles along the scan path are indicated as θ and 2θ; three detectors have $\theta = 1.5°$ and three have $\theta = 3°$.

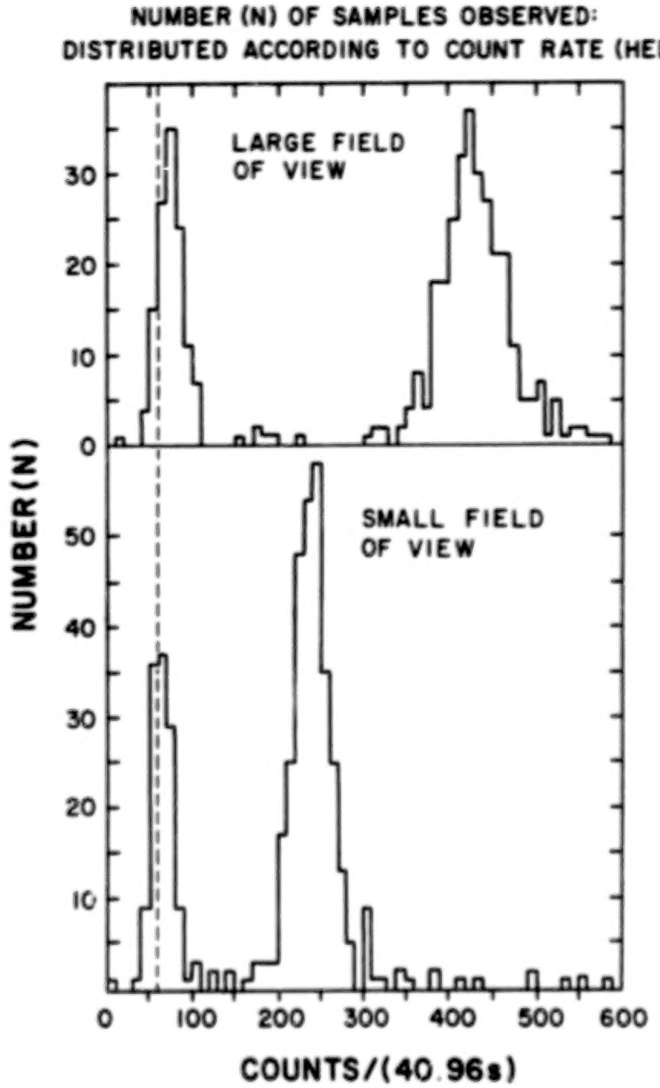


Figure 5. Histograms of sampled count rates sorted according to count per telemetry major frame (40.96 s) for HED #1 of HEAO-1 (A2), classified according to field of view (FOV)(3° × 6° for the large FOV and 3° × 3° for the small FOV).

used for the HEAO-1 (A2) experiment are no longer readily feasible. The detectors used for the HEAO-1 (A4) experiment for covering such high energy X-rays (and low energy gamma rays) involved an actively shielded layered configuration of solid scintillators (CsI(Na) and NaI(Tl)), where the light generated by one is distinguished from the other by the "phoswich" technique [Peterson, 1975]. After all the appropriate veto logic has been employed to reject many of the extraneous events, the rate of those still remaining due to causes other than the X-ray sky could be obtained by covering the detector aperture with a movable CsI(Na) scintillation crystal used in anti-coincidence with otherwise acceptable events. By accounting for the complication that the detector background depends on factors that vary in time (e.g., such as a cosmic ray flux and an induced radioactivity that vary with orbit phase) this procedure is used to estimate the absolute surface brightness of the X-ray sky at high energies [Gruber et al., 1984].

3. SPECTROSCOPY

The HEAO-1 (A2) experiment has been used to make spectral measurements of the extragalactic X-ray sky background that remains after excluding the brightest discrete sources resolved (i.e., on the order of 10 extragalactic objects per steradian). For high galactic latitudes, at $E > 3$ keV, the galactic contribution to this background is small and spectrally soft relative to the total [Iwan et al., 1982]. Employing the dual collimation scheme previously outlined, Marshall et al. [1980] determined that the spectrum of the absolute surface brightness of the extragalactic X-ray sky (3-50 keV) is well fitted by a model of optically thin thermal bremsstrahlung with $kT = 40$ keV. The scintillation detectors of the HEAO-1 (A4) experiment have been used to establish that this spectral model remains a remarkably valid description of the data up to about 100 keV [Rothschild et al., 1983; Gruber et al., 1984].

The CXB (Cosmic X-ray Background) presents us with a pronounced spectral paradox when regarded as the superposition of the contributions from discrete extragalactic sources corresponding to known classes of objects. While the majority of the extragalactic sources detected with the HEAO-2 Einstein Observatory are probably AGN (Active Galactic Nuclei) such as Seyfert galaxies and quasars [Gioia et al., 1984] the broadband continuum for the brightest

of such objects measured with HEAO-1 exhibit spectra quite different from that of the CXB. As displayed in Figure 6, these AGN exhibit power law spectra characterized by an energy spectral index value distributed in a narrow interval about $\alpha = 0.7$, apparently independent of classification or luminosity [Mushotzky, 1982]. Such a power law fails to fit the spectrum of the CXB, being clearly too steep below 10 keV and much too flat above 20 keV (see Figure 7). In general, the bright AGN observed with HEAO-1 are within the present epoch (i.e., low redshift). As shown in Figure 8, the composite broadband spectrum for a dozen of the brightest of these is well represented by a single power law up to 100 keV [Rothschild et al., 1983]. While a power-law model for the CXB spectrum is possible with $\alpha = 0.4$ over a limited band below 10 keV, it is too flat at higher energies (see Figure 7). An extrapolation of the steeper power law spectrum characterizing the contribution of unevolved Seyfert galaxies, however, could well account for essentially all of the background at energies above a few hundred keV provided there is an eventual gamma ray fall-off at several MeV [see Figure 9, taken from Rothschild et al., 1983]. Of course, the very pronounced residual CXB in the band below 100 keV of interest here remains to be explained.

Extrapolating from the sample of sources detected with the HEAO-2 Einstein Observatory, Maccacaro, Gioia, and Stocke [1984] deem it plausible to consider that most of the residual CXB could be due to quasars. Although many quasars have been observed in X-rays with HEAO-2, the data are restricted to energies below 3 keV and spectral determinations are relatively uncertain [Elvis, Wilkes, and Tananbaum, 1985]. Furthermore, even assuming that the residual CXB is indeed dominated by discrete objects (i.e., not diffuse), most such sources would be too faint to have been resolved. Hence, at this stage the spectrum characteristic of the ensemble of sources that dominate the residual CXB can only be inferred from the spectrum of the CXB itself. More precisely, we must account for that portion of the CXB arising from sources of known spectral properties (such as those observed within the present epoch) in order to isolate the residual CXB to be identified with faint discrete sources (e.g., at very large redshifts) and/or a diffuse mechanism. This has been done under a variety of approximations [De Zotti et al., 1982; Leiter and Boldt, 1982; Worrall and Marshall, 1984; Schmidt and Green, 1986], all consistent with the same general picture emerging. Specifically, the sources of the residual CXB are in this way found to be

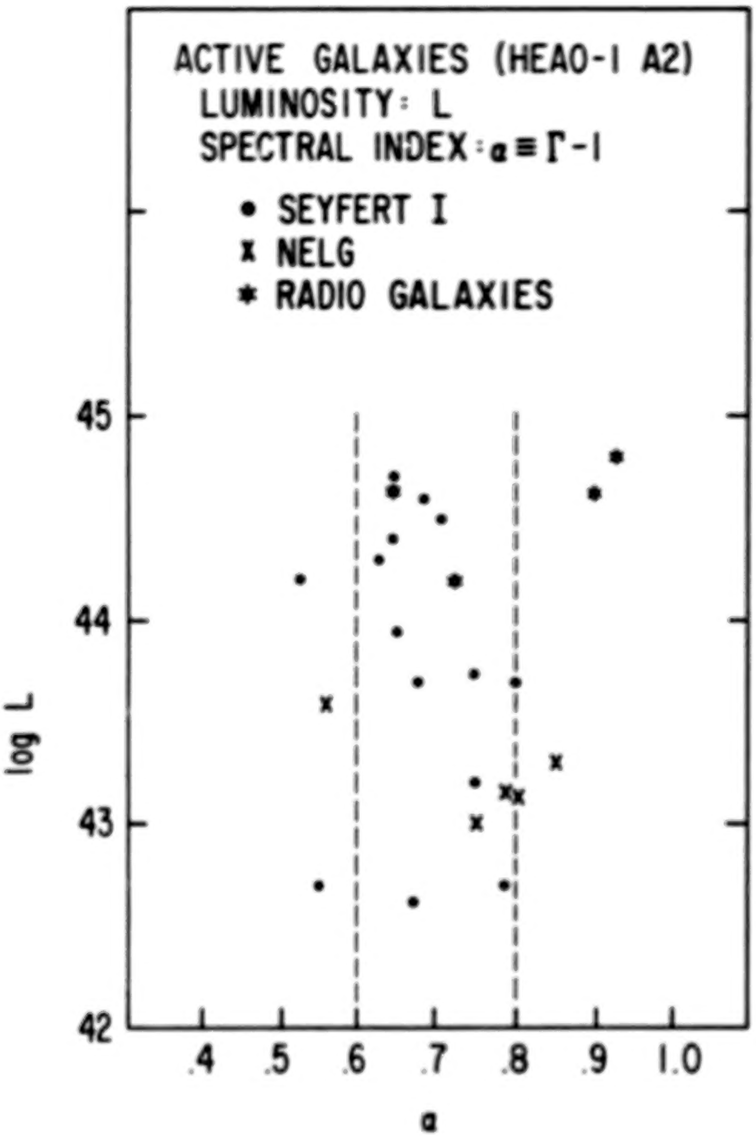


Figure 6. Active galaxies measured with the HEAO-1 (A2) experiment [Mushotzky, 1982] plotted with regard to X-ray luminosity (L ergs s^{-1}) below 20 keV and energy spectral index α.

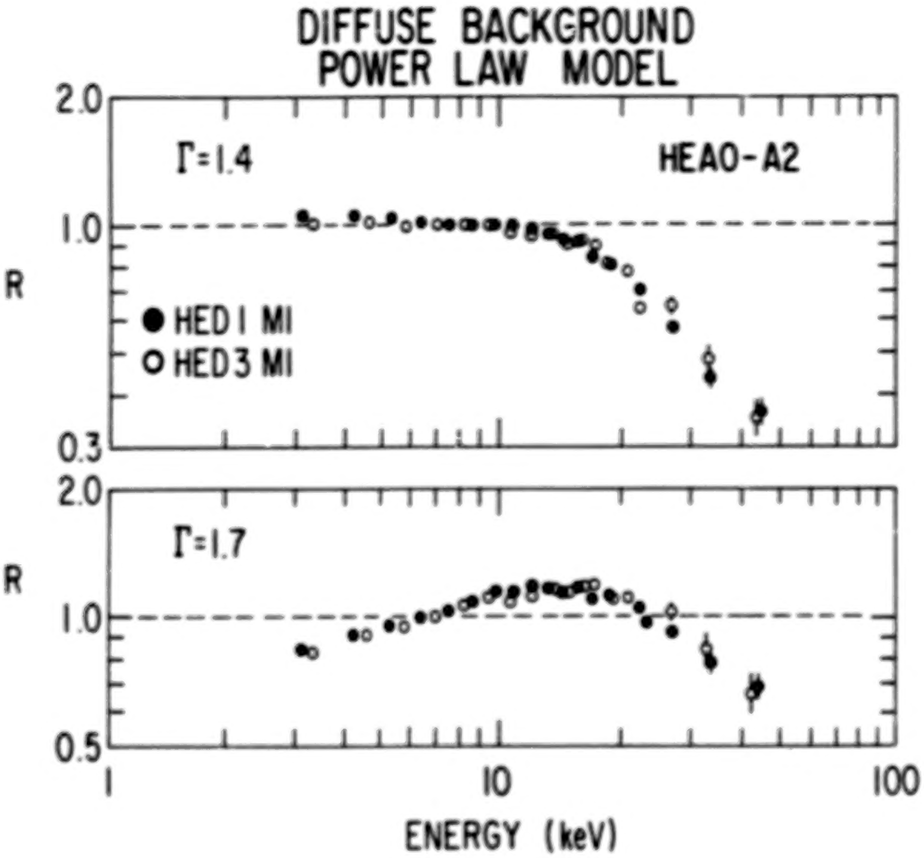


Figure 7. The ratio (R) as a function of energy of the counts observed for the X-ray background to that predicted by convolving, with the detector response function, power law spectra characterized by photon number spectral indices $\Gamma = 1.4$ and $\Gamma = 1.7$. Different symbols distinguish data obtained with HED-1 and HED-3 detectors of HEAO-1 (A2). Statistical errors are shown when larger than the size of the symbols [Marshall et al., 1980].

characterized by spectra that are extremely flat at the lowest energies and then fall off with an e-folding energy of $B(1 + z)$ in their proper frame, where $B = 23 \rightarrow 30$ keV [Leiter and Boldt, 1982] and z is redshift. For $z \gtrsim 3.5$,

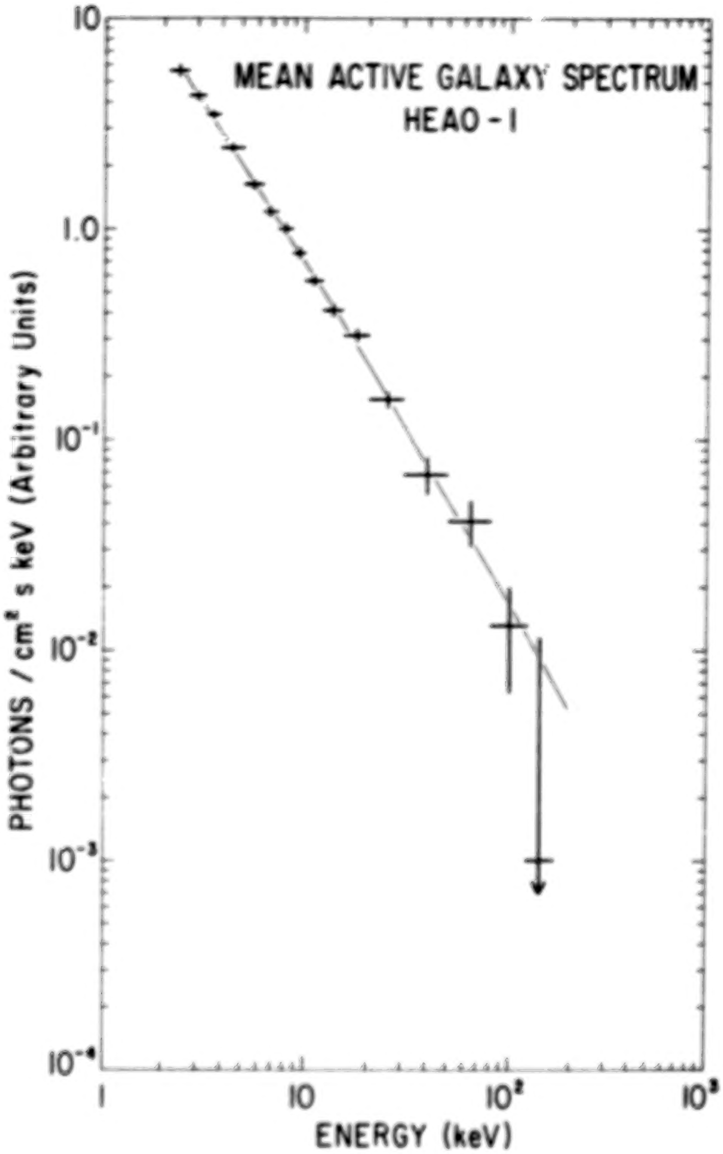


Figure 8. The mean spectrum obtained for a sample of bright AGN observed with the HEAO-1 (A2) and (A4) experiments [Rothschild et al., 1983]. The solid line is the best-fit power law model.

ORIGINAL PAGE IS OF POOR QUALITY

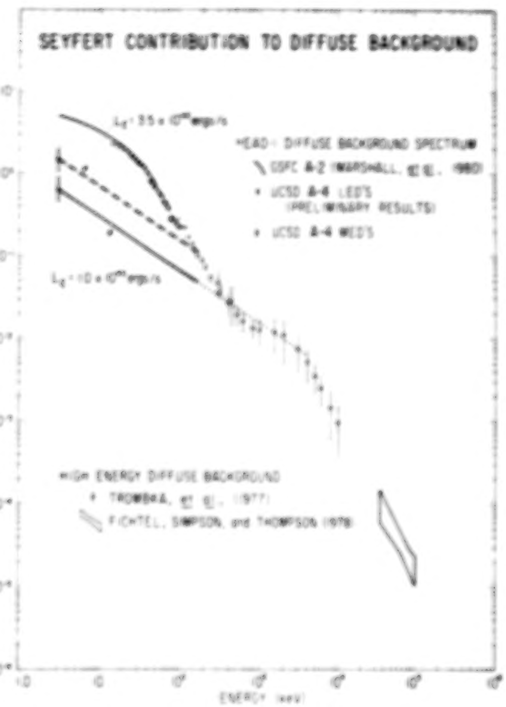


Figure 9. The Seyfert galaxy estimated contribution to the diffuse background, along with the HEAO-1 diffuse X-ray background 3-400 keV, plus points from 300 keV to 10 MeV [Trombka et al., 1977] and from 35 to 100 MeV [Fichtel, Simpson, and Thompson, 1978]. The dashed line represents the 2-165 keV power law spectra contribution from active galaxies under the assumption of a low luminosity cutoff at 3.5×10^{42} ergs s^{-1} [for a single power law luminosity function of index 2.75; see Piccinotti et al., 1982]. The solid line represents the contribution under the assumption of a break in the luminosity function index of one at 1×10^{43} ergs s^{-1}. The dotted line is an extension of the model to 4 MeV to guide the eye. The error bars at the beginning of the dashed and solid lines represent the uncertainty in the intensities due to uncertainties in the luminosity function. LED stands for the 12-165 keV Low Energy Detector, and MED stands for the 80-400 keV Medium Energy Detectors of the UCSD/MIT A-4 experiment [Rothschild et al., 1983]. The curve shown for the background (3-40 keV) is the best-fit to data obtained with the GSFC detectors of the A-2 experiment [Marshall et al., 1980].

where there is an apparent paucity of canonical quasars [Osmer, 1982; Schmidt and Green, 1983], this characteristic energy would be somewhat greater than 100 keV.

The 3-100 keV "thermal" spectrum of the overall extragalactic sky (i.e., before removing any estimated contributions from unresolved "foreground" sources) may be approximated by the following expression:

$$dI/dE = A\,[E/(3\ \mathrm{keV})]^{-\alpha} \exp(-E/B)$$

where E is photon energy (in keV), A = 5.6 keV/(keV cm^2 s sr), B = 40 keV and α = 0.29. This is to be contrasted with the broadband spectra for individually resolved sources in the present epoch. We here characterize these known discrete source spectra in terms of the same spectral form used above for the overall CXB. Accordingly, the ensemble average spectrum for clusters of galaxies may be described by an optically thin thermal model fixed by B = kT = 7 keV [Stottlemyer and Boldt, 1984], and α = 0.4; however, the contribution of such clusters is only a few percent [McKee et al., 1980]. Low redshift AGN such as Seyfert galaxies make a much more substantial contribution to the CXB (see Figure 9) and, for the brightest observed, have spectra characterized by α = 0.7 [Mushotzky, 1984] and B > 100 keV [Rothschild et al., 1983]. Considering 25 quasars brighter than visual magnitude 16 (with mean redshift z = 0.37), Worrall and Marshall [1984] have correlated data from HEAO-1 (A2) and HEAO-2 to obtain that the broadband spectrum of the ensemble may be characterized by an energy spectral index $\alpha = 0.8(+.3, -.2)$. Although the spectrum for this ensemble of quasars (mostly radio-loud) is compatible with that for Seyfert galaxies there is now some evidence from HEAO-2 that the low energy spectra (E < 3 keV) for quasars in a true optically selected sample are somewhat steeper [Elvis, 1985]. Steeper AGN spectra such as these would further aggravate the discrepancy with the CXB spectrum below 10 keV already noted for Seyfert galaxies.

Taking $\alpha = 0.6 \rightarrow 0.9$ for AGN and B = 7 keV for clusters of galaxies, the foreground spectrum arising from the composite of all these sources has been estimated [Leiter and Boldt, 1982, Appendix D]; with the normalization adopted, it amounts to 30% of the CXB over the band 3-10 keV. Subtracting this foreground from the spectrum of the overall CXB yields that of the residual

CXB which we may use for characterizing those extragalactic sources yet to be identified. As shown in Figure 10, for $E < 10$ keV this residual background is clearly flatter than that of the total CXB. For $E >> 40$ keV, we expect that Seyfert galaxies will dominate the total background [Bignami et al., 1979; Boldt, 1981; Rothschild et al., 1983], with the residual CXB becoming relatively small [see discussion of spectral model by Leiter and Boldt, 1982]. In terms of the spectral form used here the residual CXB itself must have $\alpha < 0.2$ and $B = 23 - 30$ keV. As shown in Figure 11 a good fit is obtained with $\alpha = 0$ and $B = 23$ keV. The main sources of the residual CXB should exhibit such a spectrum.

The spectrum for optically thin thermal bremsstrahlung by a hot Maxwellian plasma may be well approximated (over the band 3-100 keV) by an expression of the form used above for the CXB. In particular we note [Maxon, 1972] that $\alpha = 0.19$ for $kT = B = 200$ keV. Hence, the requirement that $\alpha < 0.2$ for the residual CXB (as viewed by the observer) implies $kT > 200$ keV at the sources of emission. Under the condition $B = 23 - 30$ keV for the observed redshifted spectrum these sources would have to be located at $z > 6$ (i.e., well beyond any known quasar). While "protogalaxies" rich in supernovae at high redshifts could well yield hot X-ray emitting galactic winds, the anticipated temperature [Bookbinder et al., 1980] might not be hot enough. Galaxy formation in an IGM (intergalactic medium) dominated by explosions, however, might give rise to a suitably hot phase [Ostriker and Cowie, 1981]. As emphasized by Fabian [1981], heating of the IGM to $kT > 200$ keV at high redshifts could provide a "natural" explanation for the spectrum of the residual CXB. According to this scenario [Guilbert and Fabian, 1986], the IGM is heated at a redshift $z < 6$ to an energy density about 40% of that in the microwave background; it initially cools mainly via Compton scattering with the microwave background (for $z > 3.5$) and subsequently cools by adiabatic expansion, decreasing in temperature to $kT < 20$ keV after a Hubble time. Between 20% and 50% of the closure density of the universe would be in this hot IGM; it would contain most of the baryons in the universe. The cosmological implications of the hot IGM hypothesis are obviously very important. While reducing the absolute magnitude of the residual CXB (i.e., increasing the level estimated for the foreground power law spectrum attributed to unresolved AGN) would mitigate the severe cosmological confrontation

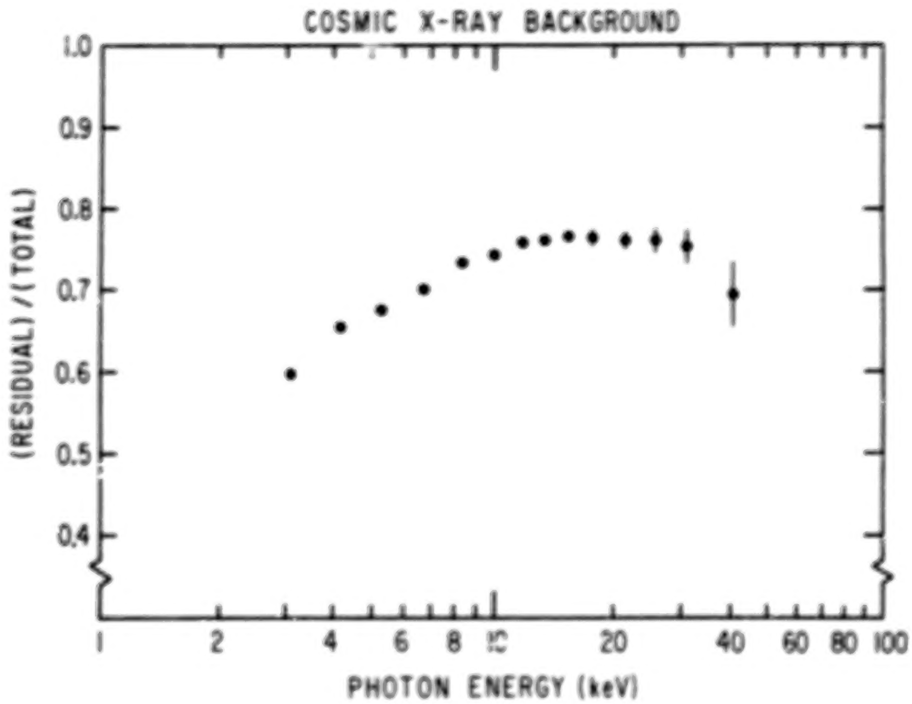


Figure 10. The ratio of the spectral density of the estimated residual CXB [Leiter and Boldt, 1982] to that of the total CXB observed with HEAO-1 (A2), plotted as a function of photon energy, in keV. Statistical errors are indicated for those data points where they exceed the size of the dots.

involved with this model, it would in fact destroy the amazingly good spectral agreement achieved with it over a broad bandwidth.

If we assume that the CXB is dominated by the integrated contributions of AGN in all their various stages of evolution, then the residual CXB spectrum may be used as a constraint on the as yet unresolved sources indicative of the youngest AGN (i.e., those of highest redshift). For example, based on

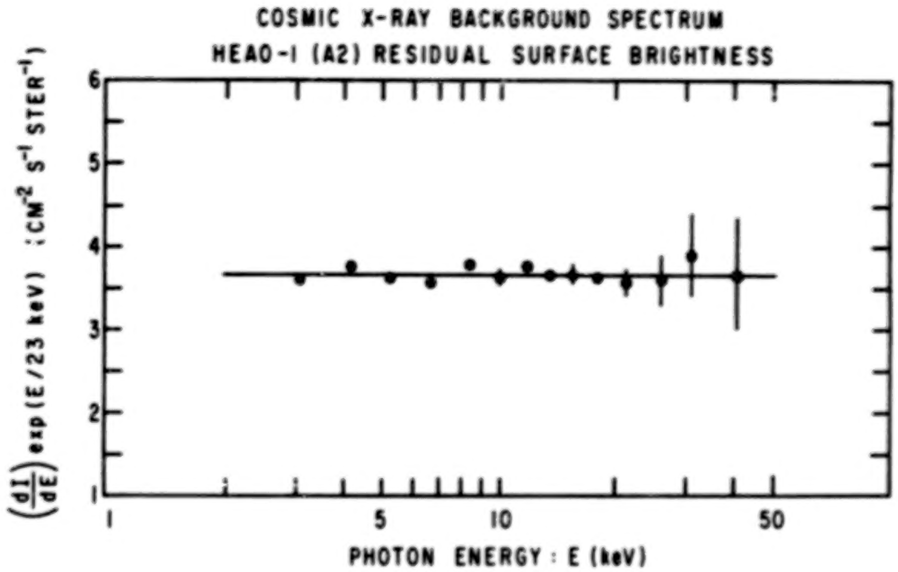


Figure 11. The residual energy spectrum for the CXB multiplied by exp(E/23 keV) as a function of photon energy [Leiter and Boldt, 1982]. In general, the statistical errors associated with the data points increase with energy and are indicated only for representative energies (~ 5 keV, 10 keV and > 20 keV).

a model of black hole disk accretion [Leiter and Boldt, 1982], a picture emerges in which:

(1) Young AGN emission occurs mainly in the X-ray band, arising from highly compact sources (i.e., a few times the gravitational radius of the central supermassive black hole in size) radiating at the Eddington luminosity limit.

(2) Canonical AGN represent later stages (less compact) exhibiting a broadband power law spectrum extending to the gamma ray regime.

Young AGN characteristic of the residual CXB could then be relatively weak optical sources [Boldt and Leiter, 1984], possibly explaining the apparent paucity of canonical quasars at $z > 3.5$. The hot, optically thin, inner disk X-ray

emission characteristic of young AGN would be diminished at a later stage, but optically thick radiation from a cooler disk may explain the "UV bump" observed for AGN [Malkan and Sargent, 1982]. In a similar sense, the residual CXB that appears above the broadband nonthermal continuum spectrum of the overall background (see Figure 1) may be viewed as indicative of a "thermal bump" dominating the emission of *young* AGN.

In our discussions interpreting the residual CXB we have implicitly assumed that the background observed is the superposition of radiation coming directly from all the sources of emission in the universe. This is the kind of arithmetic prescribed by any standard cosmology which excludes the possibility of a cosmological albedo due to the entire universe in which all the sources are embedded. With Friedmann cosmology, which is not temporally homogeneous, once radiation is emitted it propagates away never to return. In any temporally homogeneous model based on a closed space, however, the radiation may circulate many times around before it is scattered or absorbed [Segal, 1985]. As viewed by Segal [1983] the microwave background could thereby be intrinsically diffuse cosmological albedo arising from dust. If such is the case, then the cosmological albedo responsible for the residual CXB would probably be due to Compton scattering from gas and plasma. Because of radiative transfer effects associated with Compton scattering such an albedo would occur mainly at energies less than mc^2, explaining the possible absence of any significant amount of genuinely diffuse background above a few hundred keV.

4. ISOTROPY

When viewed over large angular scales at high galactic latitudes the CXB is essentially isotropic [Boldt, 1981; Shafer, 1983; Shafer and Fabian, 1983]. A small global anisotropy observed in the overall background (> 3 keV) is due mainly to our galaxy. This galactic component accounts for about 2% of the background at high latitudes; characterized by kT = 9 keV its spectrum is softer than that of the CXB [Iwan et al., 1982]. The observed small-scale fluctuations of CXB surface brightness may be ascribed to statistical variations in the population of unresolved discrete sources [Shafer, 1983] among the pixels examined. The number-flux relation ($N \propto S^{-1.5}$) for the bright extragalactic X-ray sources (> 3 keV) observed with HEAO-1 [Piccinotti

et al., 1982] may be extrapolated to the regime of unresolved faint sources needed for evaluating such fluctuations. Assuming typical source spectra, this extrapolation is consistent with the number-flux relation for faint sources directly obtained [Gioia et al., 1984] from HEAO-2 observations (< 3 keV). The apparent surface brightness fluctuations arising from those sources in this population unresolved with HEAO-1 are found to be statistically consistent with the variations observed [Shafer, 1983]. The upper limit to any residual variations implies that sources fainter than those already detected in the HEAO-2 deep survey [Giacconi et al., 1979b] can make only a relatively small contribution to the total fluctuations. Assuming an energy spectral index $\alpha = 0.7$, the flux threshold S_0(1-3 keV) = 2.6×10^{-14} ergs cm^{-2} s^{-1} for the source sample obtained in this deep survey corresponds to S_0(3-10 keV) = 4.0×10^{-14} ergs cm^{-2} s^{-1}. With this scaling, the population of sources represented by the HEAO-2 sample accounts for only 18(+7, −7)% of the CXB (3-10 keV). Hence, it is evident that the CXB flux itself must be dominated by sources other than those responsible for most of the fluctuations. This means that the number of sources fainter than the HEAO-2 deep survey limit must be relatively high [Shafer, 1983] or that the CXB is largely diffuse.

Models for the origin of the CXB attribute it to numerous, weak, unresolved discrete sources (e.g., low luminosity AGN, as discussed by Elvis, Soltan, and Keel, 1984), to some diffuse mechanism (e.g., thermal bremsstrahlung from a hot IGM), or to an unresolved population of discrete sources which evolve substantially with redshift (e.g., quasars). Turner and Geller [1980] have shown that comparison of the CXB surface brightness variations with flux variations in the integrated light from galaxies can measure the relative contributions of these possible sources of the CXB. Applying the technique to Uhuru data, they find that the absence of correlation between the optical flux variations and the X-ray flux variations sets an upper limit of about 50% on the fraction of the CXB originating with any classes of X-ray sources substantially represented among bright ($m_{pg} < 15.5$) galaxies. The data are consistent with nearly all of the CXB being due to diffuse emission or to a class of sources whose density and/or luminosity increases rapidly with redshift. Further studies of such correlations (or limits on them) are being carried out with HEAO-1 (A2) data on the CXB [Persic et al., 1986], including a direct

comparison with the infrared sky as observed with the IRAS (Infrared Astronomy Satellite).

Since the HEAO-1 mission surveyed the sky with scan paths that follow great circles which always traverse the ecliptic poles, the most straightforward and reliable way to investigate possible weak global anisotropies of the CXB involves referring data to ecliptic coordinates. Figure 12 shows the geometry to be considered, expressed in ecliptic coordinates; latitude is β, longitude is λ. In this representation, supplementary longitudes λ and $(\lambda + 180°)$ are used to identify the dual traces of any given great circle scan path. The loci of galactic and supergalactic equator crossings are shown by the curves labeled as such. Some specific directions of particular interest cluster together in the general vicinity of $\lambda = 180°$ and are separately identified by numbers, as follows:

#1 (at $\lambda = 185°$) gives the direction of the dipole anisotropy characteristic of the microwave background, as measured by Smoot, Gorenstein, and Muller [1977].

#2 (at $\lambda = 164°$) gives the direction of the microwave dipole anisotropy, as measured by Cheng et al. [1979].

#3 (at $\lambda = 184°$) gives the direction of the Virgo Cluster near the center of the Supergalaxy [de Vaucouleurs, 1958].

#4 (at $\lambda = 170°$) is the solar velocity direction relative to distant galaxies [Hart and Davies, 1982] and, within errors, is consistent with the preferred direction for the microwave background.

The preferred direction to be associated with the microwave background is close to the ecliptic equator, which makes the investigation of correlations particularly well suited to studies carried out with HEAO-1. To minimize galactic effects (i. e., by avoiding appreciable variations of galactic latitude relative to this direction) only those data corresponding to a band within 24° of the ecliptic equator are considered here; this band is outlined on Figure 12 with dashed lines. Even so, unavoidable galactic effects do become important when considering longitudes close to where the galactic equator crosses the ecliptic

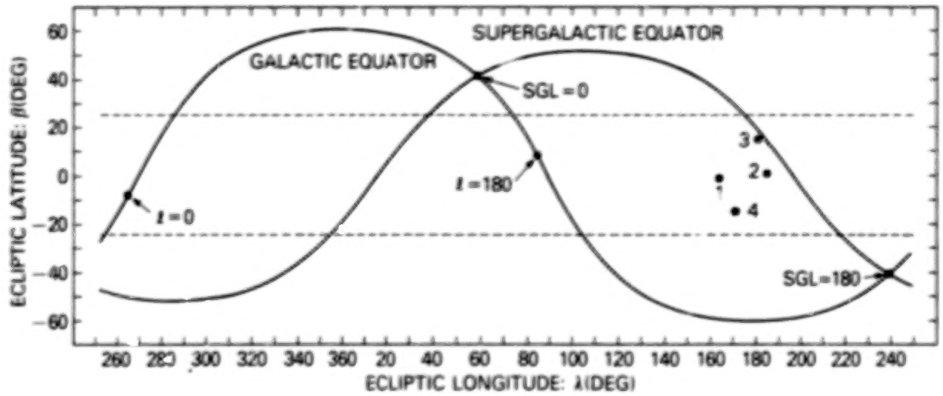


Figure 12. Special directions on the celestial sphere, expressed via ecliptic coordinates, in degrees (latitude: β, longitude: λ). The galactic and supergalactic equators are indicated by solid curves. The galactic center and anti-center are denoted by galactic longitudes $l = 0$ and $l = 180°$, respectively Reference supergalactic longitudes are donated by SGL = 0 and SGL = 180°. The dashed lines outline the band within 24° of the ecliptic plane used for evaluating anisotropies (see text). #1 (λ = 185°) and #2 (λ = 164°) indicate the direction of the dipole anisotropy of the microwave background as determined by Smoot et al. [1977] and Cheng et al. [1979], respectively #3 (λ = 184°) is the direction of the Virgo Cluster and #4 (λ = 170°) is the direction of the solar velocity relative to distant galaxies [Hart and Davies, 1982].

plane (i.e., at λ = 89° and λ = 269°). Excluding the contribution of resolved sources, the average surface brightness for the band within 24° of the ecliptic equator has been determined as a function of ecliptic longitude. Deviations from isotropy so obtained are represented in Figure 13.

The circle shown in Figure 13 represents isotropy. Percentage deviations up to about one percent are here plotted as a function of ecliptic longitude (λ). Each interval corresponds to a region of about 8×10^2 square degrees for which the average surface brightness is determined to a statistical uncertainty (one sigma) of 0.1% to 0.2%, arising from photon counting noise; the root

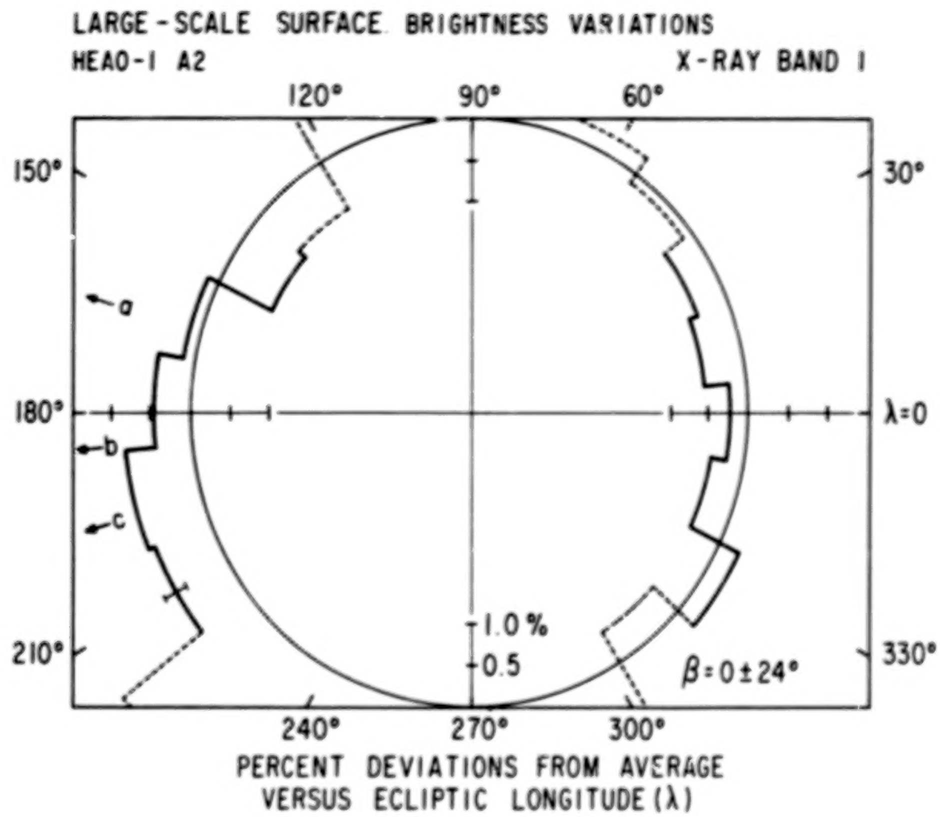


Figure 13. Percentage deviations from isotropy (represented by the circle) for the average surface brightness near the ecliptic plane ($\beta = -24° \rightarrow +24°$) as a function of ecliptic longitude (λ). Data obtained with HED units of HEAO-1 (A2) in all-sky survey. The one-sigma error bar shown corresponds to photon counting statistics. Special directions indicated are for the microwave dipole anisotropy results of Smoot et al. [1977] denoted by (a) and Cheng et al. [1979] denoted by (b), and the longitude where the supergalactic plane crosses the ecliptic equator (c). Surface brightness deviations for those ecliptic longitudes close to the galactic equator (i.e., near $\lambda = 89°$ and $\lambda = 269°$) are indicated with dashed lines.

mean squared (rms) fluctuation in apparent surface brightness among such intervals, due only to unresolved faint sources, is estimated to be about 0.3%. The galactic plane crosses the ecliptic equator at $\lambda = 89°$ and $\lambda = 269°$; those deviations in isotropy expected to be influenced most by this are shown as dashed lines. In this representation, the preferred directions of interest are:

(a) the dipole anisotropy of the microwave background as determined by Smoot et al. [1977]

(b) the dipole anisotropy of the microwave background as determined by Cheng et al. [1979]

(c) the longitude where the plane of the Supergalaxy crosses the ecliptic equator

Considering only the solid portion of this plot, where galactic effects should be minimal, a weak residual large-scale asymmetry of the CXB becomes evident, one that repeats for independent sky surveys carried out with HEAO-1 at six-month centers.

Within estimated errors, the motion of the Sun relative to the backdrop of galaxies extending out to 70 Mpc [Hart and Davies, 1982] is consistent with the velocity vector inferred from a Compton-Getting [1935] interpretation of the dipole anisotropy observed for the cosmic microwave background. Based upon data from the HEAO-1 A2 cosmic X-ray experiment, the large-scale anisotropy of the CXB also appears to be compatible with such an interpretation [Shafer, 1983; Shafer and Fabian, 1983]. Considering the relatively large uncertainties involved, however, the anisotropy of the CXB is also consistent with an interpretation based on a possible component of the extragalactic background correlated with a direction defined by the physical center of the Supergalaxy [Boldt, 1981; Shafer, 1983], at the supergalactic longitude (104°) associated with the Virgo Cluster. Furthermore, the amplitude of large-scale variations in the X-ray background associated with the Milky Way at high galactic latitudes might not be small in magnitude relative to the anisotropy attributed to the extragalactic background [Iwan et al., 1982]. In order to obtain a diagnostic of these possible complications, we postulate that the dipole

anisotropy of the cosmic microwave background provides a precise indication of the direction of the observer's velocity relative to the proper (co-moving) frame of the CXB and examine the energy dependence of the fore-aft asymmetry of the X-ray background relative to this direction as a reference (see Figure 14).

The fore-aft asymmetry arising from the observer's velocity v/c (in units of the velocity of light) relative to the proper frame of an isotropic background of electromagnetic radiation having a spectrum of the form used here for the CXB is given by:

$$\Delta I/I = (v/c)[1 + \cos(\delta)]\ [3 + \alpha + (E/B)]$$

where δ is the half-cone angle defining the two opposite regions of the sky relative to the axis of motion; for the total CXB, $\alpha = 0.29$ and $B = 40$ keV. It is important to note that this Compton-Getting asymmetry increases with photon energy, becoming most pronounced for $E > B$. On the other hand, if there is a spatial inhomogeneity in the emission responsible for a background component characterized by a spectrum that is softer than the overall background (e.g., thermal emission with $kT < 40$ keV), then the corresponding asymmetry in the background flux would decrease with photon energy. In fact, we already know of two such X-ray background components characterized by $kT < 40$ keV. In particular, the unresolved galactic emission ($E > 3$ keV) at high latitudes may be described with $kT = 9$ keV [Iwan et al., 1982] and the background component due to clusters of galaxies may be characterized by $kT = 7$ keV [Stottlemyer and Boldt, 1984].

Corresponding to Figure 14 the ratio (R) of the fore-aft asymmetry for the band (10.3 − 20) keV to that for the band below 10.3 keV is given by $R = 1.72(+0.40, -0.40)$. This value for R (considering two-sigma statistical uncertainty) is compatible with the Compton-Getting effect as an explanation for the observed asymmetry (see Figure 14) but is incompatible with an explanation based on any anisotropic thermal background components having $kT \leqslant 9$ keV. However, an anisotropy in the foreground distribution of faint

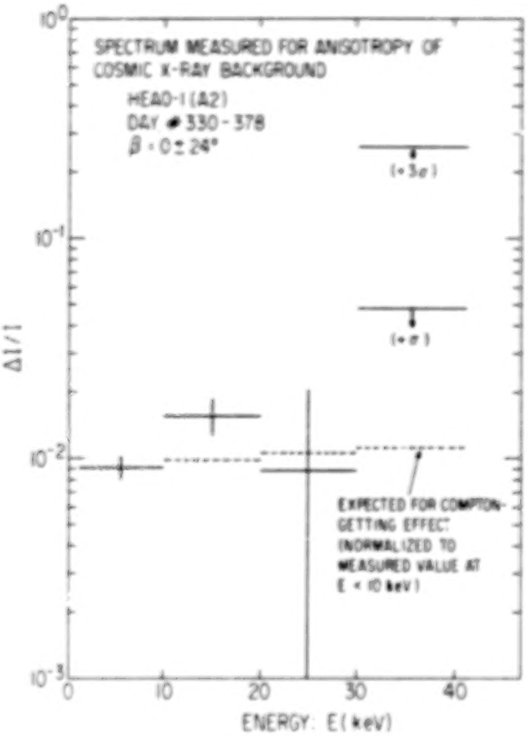


Figure 14. The fore-aft asymmetry (ΔI/I) of the cosmic X-ray background, as measured with the HEAO-1 (A2) experiment, is exhibited as a function of photon energy E(keV). The reference direction is that defined by ecliptic coordinates: β (latitude) = 0° and λ (longitude) = 178°, corresponding to δ (declination) = 1° and a (right ascension) = 178°. The average surface brightness for a region 48° × 48° (Δβ × Δλ), centered on the reference direction, is compared with that for an equivalent region in the opposite direction. Resolved sources, brighter than 2.1 × 10^{-11} ergs cm^{-2} s^{-1} (2-10 keV), have been excluded in the sample analyzed. Data were obtained with xenon gas proportional counters (high energy detectors HED-I and HED-III) and an argon counter (MED: medium energy detector) during the 48-day interval when great-circle scans with HEAO-1 covered the regions of the sky considered here. The one-sigma errors shown correspond to photon counting statistics. Dashed lines indicate the asymmetry expected for the Compton-Getting effect (normalized to the value measured for the band 1.3-10.3 keV).

unresolved extragalactic sources (e.g., correlated with the Supergalaxy), exhibiting spectral indices $\alpha \leqslant 0.7$, can not now be ruled out as a major cause of the asymmetry observed for the CXB.

The absolute precision of the CXB asymmetry measured with HEAO-1 (A2) below about 20 keV is limited by fluctuations in apparent surface brightness arising from unresolved faint foreground sources and not by the number of photons detected (i.e., this is the case for the two lowest energy bins exhibited in Figure 14). If the surface brightness measured could be isolated to those small pixels (i.e., of arc-minute size) devoid of foreground sources [down to a level corresponding to the limiting sensitivity S_o of the HEAO-2 Einstein Observatory; Giacconi et al., 1979b], then the large-scale CXB anisotropy could be determined to a precision comparable with that for the cosmic microwave background. In this way, the CXB could be used for determining the vector velocity of the co-moving frame in which to examine anisotropies *intrinsic* to the cosmic microwave background, an issue that is central to the observational basis of modern cosmology.

5. OUTLOOK

The residual CXB needs to be measured directly; this will involve resolving out discrete sources in the background. The magnitude and fluctuations expected for the integral contribution of these individual sources to the CXB at any given energy band may be obtained from the corresponding log(N)-log(S) relation that describes the counts of such sources. It is usual to characterize this relation by an index γ, generally a weak function of S, defined by

$$\gamma(S) = -d[\log(dN/dS)]/d[\log(S)].$$

For Euclidean space $\gamma = 2.5$. For the case of non-evolving sources (i.e., constant co-moving density and invariant source characteristics) standard cosmological models have $\gamma = 2.5$ only for the brightest (nearest) sources, with γ becoming smaller as S decreases. Since the extragalactic foreground sources detected with HEAO-2 exhibit a log(N)-log(S) relation consistent with

γ = 2.5 (i.e., N $\propto S^{-1.5}$) all the way down to the Einstein Observatory's survey limit (S_o), we must consider the effects of evolution. Limiting the HEAO-2 sample to AGN alone yields γ = 2.71(+.15, −.15), indicating that this evolution is probably substantial [Gioia et al., 1984]. By contrast, the number-flux relation for the subsample made up of clusters of galaxies is characterized by γ = 2.04(+.23, −.23), less than the Euclidean value (as anticipated).

The AXAF (Advanced X-ray Astrophysics Facility) to be co-orbiting with the Space Station will have the capability of resolving hard X-ray sources (3-10 keV) contributing to the CXB. The sensitivity for detecting point sources with the AXAF focusing X-ray telescope is expected to be one to two orders of magnitude better than that for the HEAO-2 Einstein telescope [Zombeck and Ramage, 1983]. The fraction of the CXB yet to be resolved into discrete objects depends on the source number-flux relation at $S < S_o$; power law extrapolations using γ = 2.5(+0.5, −0.5) are exhibited in Figure 15 for purposes of estimation. Assuming that the CXB point-source contribution to be resolved with the AXAF telescope arises mainly from AGN with canonical spectra (α = 0.7) and that the number-flux relation persists to about $0.1S_o$ with $\gamma \geq 2$, we note that the resolved portion would be in excess of 35% (for the band 3-10 keV). Since the 3-10 keV spectrum of the CXB is significantly flatter than that of such AGN, however, their contribution over this band could not be this large [De Zotti et al., 1982]. As exhibited in Figure 15, the number of such resolved sources is expected to be much less than one per square arc-minute pixel.

Based on data from the imaging proportional counter in the focal plane of the Einstein Observatory telescope, Hamilton and Helfand [1986] have already investigated the surface brightness of the X-ray background (1-3 keV) with arc-minute resolution, thereby obtaining upper limits on small-scale fluctuations which suggest that the residual CXB could very well be predominantly diffuse. If due to point sources, they find that the number needed would have to exceed 5×10^3 degree^{-2}, much larger than the estimated total number of quasars [Schmidt and Green, 1986]. Assuming that these sources are at $z \leq 4$ (with $\Omega = 1$) the residual CXB spectrum (see Figure 11) implies that their average X-ray luminosity would have to be less than 10^{45} erg s^{-1}. If these objects are precursor AGN whose thermal X-radiation is at a level close to the

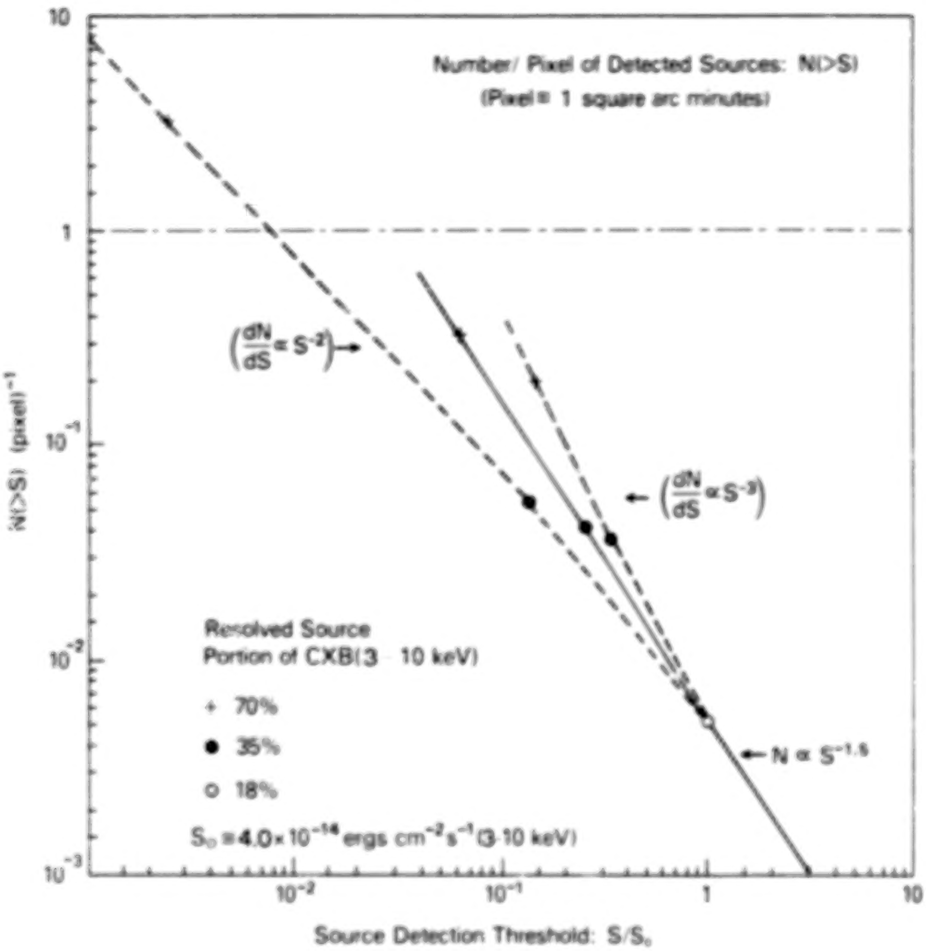


Figure 15. Extrapolation of source counts to $S/S_o < 1$, normalized to HEAO-2 measurements at $S/S_o > 1$ [Gioia et al., 1984]. Points designated by various symbols indicate corresponding percentage of the CXB resolved (3-10 keV), shown separately for γ = 2, 2.5 and 3 (see text). For an energy spectral; index α = 0.7, S_o (3-10 keV) = 4.0×10^{-14} ergs cm^{-2} s^{-1} corresponds to S_o(1-3 keV) = 2.6×10^{-14} ergs cm^{-2} s^{-1} of the HEAO-2 deep survey [Giacconi et al., 1979b]. N(>S), average number of sources per pixel brighter than S, is plotted versus S/S_o. Pixel size is 1 square arc-minute.

Eddington luminosity limit (see Section 3), then the characteristic central compact mass involved has to be less than that measured for typical Seyfert galaxies in the present epoch [Wandel and Mushotzky, 1986], as expected.

Considering that a fraction (f) of the residual CXB over the band 3-10 keV (having the spectrum indicated in Figure 11) arises from a presently unknown population of faint discrete hard X-ray sources, their average flux would be given by

$$\langle S(3\ \text{keV-}10\ \text{keV})\rangle = 2.6\times10^{-15}(f/N)\ \text{ergs cm}^{-2}\ \text{s}^{-1}$$

where N is the number of such objects per square arc-minute. For the average source in this population to be detectable with AXAF would require $N < (4f)$ per square arc-minute. This overall performance level is probably sufficient for the study of that particular candidate population of young precursor AGN sources already postulated for the residual CXB [Leiter and Boldt, 1982]. The portion (1-f) of the residual CXB unresolved with AXAF (e.g., due to a diffuse component) would yield a photon flux density at the focal plane of the telescope equal to $1.4\times10^{-5}(1\text{-}f)\ \text{mm}^{-2}\ \text{s}^{-1}$ for the band 3-10 keV. Since the internal background for imaging X-ray detectors over this energy band is generally on the order of $10^{-5}\ \text{mm}^{-2}\ \text{s}^{-1}$ or greater, unresolved and/or diffuse components of the residual CXB may not be very well suited for study with the AXAF telescope. "Faster" optics are desirable for such weak surface brightness investigations, but the angular resolution might not need to be better than an arc-minute for isolating those pixels to be studied which are devoid of foreground point sources (see Figure 15).

If the residual CXB is dominated by point sources having $S > 0.01S_0$ then they will be detectable with AXAF and their average spectrum can be determined. As shown by Worrall and Marshall [1984], based on HEAO-1 data, the residual CXB spectrum over the band 3-10 keV is clearly incompatible with $\alpha > 0.68$ (i.e., the 90% confidence lower limits to the spectra index characterizing their quasar sample). Over this band suitable to AXAF, the sources of the residual CXB should exhibit an average spectrum with $\alpha = 0.29$ (see Figure 16), indicating a population of objects that are distinct from

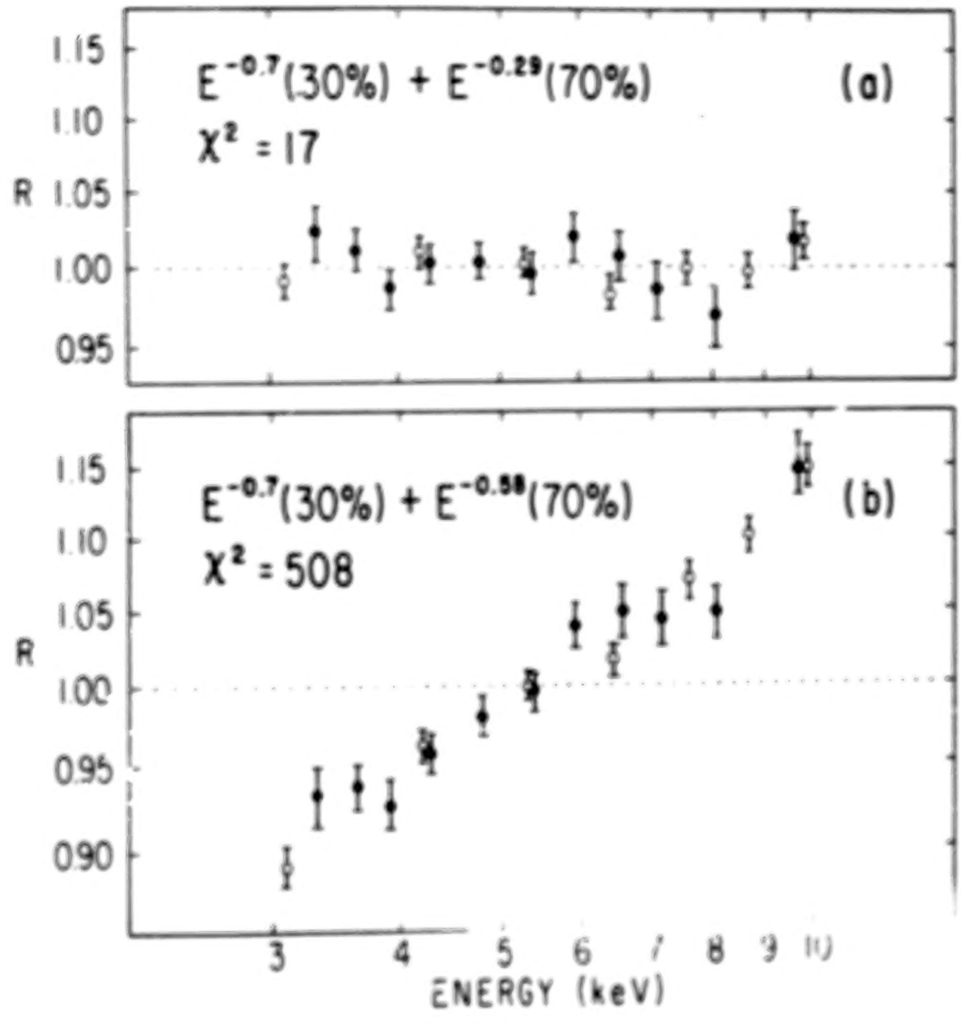


Figure 16. The ratio (R) of observed counts for the X-ray background to those predicted (3-10 keV) for combinations of incident power law spectra [Worrall and Marshall, 1984]. Open circles are measurements with the HEAO-1 (A2) HED-1 detector. Closed circles are with the MED.

the AGN X-ray sources already studied (see Figure 6). This would be consistent with the conclusion reached by Narlikar and Burbidge [1983] who, considering the upper limit to the extragalactic night sky background light, find it to be unlikely that canonical quasar-like objects dominate the CXB (i.e., the ratio of X-ray to optical emission must be extraordinarily large for the principal sources of the CXB).

If much of the residual CXB is due to a hot IGM, then there would be spatial structure to this sky background on a scale defined by the clumping of the gas. In this connection we note that the large-scale structures associated with galaxy clustering in recent epochs (e.g., superclusters, giant voids) may be remnant indicators of the distribution of relatively young matter at the epoch of galaxy formation [Oort, 1983]. Assuming a present-epoch scale $\simeq 0.04$ (c/H_0), as described by Bahcall and Burgett [1986], the related structure at any epoch would appear to us with an angular scale greater than the field of view ($\sim$ one degree) of the AXAF telescope.

Direct measurements of the residual CXB over a bandwidth sufficiently large to cover the spectrum (i.e., $\Delta E > B$) are beyond the capability of already developed focusing X-ray optics. The sensitive all-sky survey of soft X-rays (< 2 keV) to be provided by the focusing X-ray telescope of the forthcoming Roentgen Satellite (ROSAT) [Trumper, 1984] is likely to be complicated by galactic emission. Although the precise direction to be associated with the dipole anisotropy of the residual CXB (up to 10 keV) could be obtained via an all-sky scan with a broadband "fast" focusing X-ray telescope [such as developed by Serlemitsos et al., 1984], a verification of the Compton-Getting velocity interpretation would still require examining how the anisotropy varies at substantially higher energies.

REFERENCES

Alexander, J., and Clark, T., 1974, in *High Energy Particles and Quanta in Astrophysics*, ed. F. McDonald and C. Fichtel (Cambridge: MIT Press), pp. 299-300.

Bahcall, N., and Burgett, W., 1986, *Astrophys. J.*, **300**, L35.

Bignami, G., Fichtel, C., Hartman, R., and Thompson, D., 1979, *Astrophys. J.*, **232**, 649.

Boldt, E., 1971, *Bull APS*, **224**, 534.

Boldt, E., 1974, in *High Energy Particles and Quanta in Astrophysics*, ed. F. McDonald and C. Fichtel (Cambridge: MIT Press), pp. 401-406.

Boldt, E., 1981, *Comments Ap.*, **9**, 97.

Boldt, E. et al., 1979, in (COSPAR) *X-ray Astronomy*, ed. W. Baity and L. Peterson (Oxford: Pergamon Press), p. 433.

Boldt, E., and Leiter, D., 1984, *Astrophys. J.*, **276**, 472.

Bookbinder, J. et al., 1980, *Astrophys. J.*, **237**, 647.

Cheng, E. et al., 1979, *Astrophys. J.*, **232**, L139.

Clark, T., Brown, L., and Alexander, J., 1970, *Nature*, **228**, 847.

Code, A., and Welch, G., 1982, *Astrophys. J.*, **256**, 1.

Compton, A., and Getting, I., 1935, *Phys. Rev.*, **47**, 817.

de Vaucouleurs, G., 1958, *Astrophys. J.*, **63**, 253.

De Zotti, G. 1982, *Acta. Cosmol.*, **11**, 65, "Second International Krakow School of Cosmology", Jablonna, Poland.

De Zotti, G. et al., 1982, *Astrophys. J.*, **253**, 47.

Dube, R., Wickes, W., and Wilkinson, T., 1979, *Astrophys. J.*, **232**, 333.

Elvis, M., 1985, in Proc. Japan/U.S. Symp. (Jan. 1985), eds. Y. Tanaka and W. Lewin (Tokyo: Inst. Space Astron. Sc.), p. 291; CFA Preprint #2126.

Elvis, M., Wilkes, B., and Tananbaum, H., 1985, *Astrophys. J.*, **292**, 357.

Elvis, M., Soltan, A., and Keel, W., 1984, *Astrophys. J.*, **283**, 479.

Fabian, A., 1981, Proc. Tenth Texas Symp. Rel. Astr., *Ann. N.Y. Acad. Sci.*, **375**, 235.

Fichtel, C., Simpson, G., and Thompson, D., 1978, *Astrophys. J.*, **222**, 833.

Fichtel, C., and Trombka, J., 1981, *Gamma Ray Astrophysics*, NASA SP-453.

Garmire, G., 1979, in *HEAO Science Symposium*, ed. C. Dailey and W. Johnson, NASA CP-2113, p. 49.

Giacconi, R., 1986, this volume.

Giacconi, R., Gursky, H., Paolini, F., and Rossi, B., 1962, *Phys. Rev. Letters*, **9**, 439.

Giacconi, R., Gursky, H., and van Speybroeck, L., 1968, *Ann. Rev. Astron. Astrophys.*, **6**, 373.

Giacconi, R. et al., 1979a, *Astrophys. J.*, **230**, 540.

Giacconi, R. et al., 1979b, *Astrophys. J.*, **234**, L1.

Gioia, I. et al., 1984, *Astrophys. J.*, **283**, 495.

Gruber, D., Rothschild, R., Matteson, J., and Kinzer, R., 1984, in *Conf. Proc. X-ray and UV Emission From Active Galactic Nuclei*, ed. W. Brinkman and J. Trumper (Garching: Max Planck Institute), p. 129.

Guilbert, P., and Fabian, A., 1986, *MNRAS*, **220**, 439.

Hamilton, T., and Helfand, D., 1986, *Astrophys. J.*, submitted.

Hart, L., and Davies, R., 1982, *Nature*, **297**, 191.

Holt, S., 1974, ed. Proc. GSFC Workshop Electron Contam. X-ray Astron. Exp., GSFC X-661-74-130.

Iwan, D. et al., 1982, *Astrophys. J.*, **260**, 111.

Joubert, M. et al., 1983, *Astron. Astrophys.*, **128**, 114.

Leiter, D., and Boldt, E., 1982, *Astrophys. J.*, **260**, 1.

Maccacaro, D., Gioia, I., and Stocke, J., 1984, *Astrophys. J.*, **283**, 486.

Malkan, M., and Sargent, W., 1982, *Astrophys. J.*, **254**, 22.

Marshall, F. et al., 1980, *Astrophys. J.*, **235**, 4.

Maxon, S., 1972, *Phys. Rev. A.*, **5**, 1630.

McCammon, D. et al., 1983, *Astrophys. J.*, **269**, 107.

McKee, J. et al., 1980, *Astrophys. J.*, **242**, 843.

Mushotzky, R., 1982, *Astrophys. J.*, **256**, 92.

Mushotzky, R., 1984, *Adv. Space Res.*, **3**, 157.

Narlikar, J., and Burbidge, G., 1983, *Astron. Astrophys.*, **118**, 154.

Oort, S., 1983, *Ann. Rev. Astron. Astrophys.*, **21**, 373.

Osmer, P., 1982, *Astrophys. J.*, **253**, 28.

Ostriker, J., and Cowie, L., 1981, *Astrophys. J.*, **243**, L127.

Paresce, F., McKee, C., and Bowyer, S., 1980, *Astrophys. J.*, **240**, 387.

Persic, M. et al. 1986, in preparation.

Peterson, L., 1975, *Ann. Rev. Astron. Astrophys.*, **13**, 423.

Piccinotti, G. et al., 1982, *Astrophys. J.*, **253**, 485.

Razin, V., 1960, *Izv. Vysshikh Uchebn. Zacedenii, Radiofiz.*, **3**, 584.

Rothschild, R. et al., 1979, *Sp. Sci. Instr.*, **4**, 269.

Rothschild, R. et al., 1983, *Astrophys. J.*, **269**, 423.

Schmidt, M., and Green, R., 1983, *Astrophys. J.*, **269**, 352.

Schmidt, M., and Green, R., 1986, *Astrophys. J.*, **305**, 68.

Segal, I. E., 1983, *Phys. Rev. D.*, **28**, 2393.

Segal, I. E., 1985, in *Cosmic Background Radiation and Fundamental Physics*, Proc. Third Rome Meeting on Astrophysics, ed. F. Melchiorri (Bologna: Italian Physical Society), p. 209.

Serlemitsos, P. et al., 1984, *IEEE Trans. Nucl. Sci.*, **NS-31**, 786.

Shafer, R., 1983, Ph.D. dissertation, University of Maryland, NASA TM 85029.

Shafer, R., and Fabian, A., 1983, in *Early Evolution of the Universe and its Present Structure*, ed. G. Abell and G. Chincarini (IAU), **104**, 333.

Simon, A., 1977, in *Radio Astronomy and Cosmology*, ed. D. Jauncey (Dordrecht: D. Reidel Publishing Co.), p. 349.

Smoot, G., Gorenstein, M., and Muller, R., 1977, *Phys. Rev. Letters*, **39**, 898.

Stottlemyer, A., and Boldt, E., 1984, *Astrophys. J.*, **279**, 511.

Toller, G., 1983, *Astrophys. J.*, **266**, L79.

Trombka, J. et al., 1977, *Astrophys. J.*, **212**, 925.

Trumper, J., 1984, *Phys. Scripta*, **T7**, 209.

Tucker, W., 1984, *The Star Splitters*, NASA SP 466.

Turner, E., and Geller, M., 1980, *Astrophys. J.*, **236**, 1.

Wandel, A., and Mushotzky, R., 1986, *Astrophys. J.*, **306**, L61.

Weiss, R., 1980, *Ann. Rev. Astron. Astrophys.*, **18**, 489.

Weller, C., 1983, *Astrophys. J.*, **268**, 899.

Worrall, D., and Marshall, F., 1984, *Astrophys. J.*, **276**, 434.

Zombeck, M., and Ramage, S., 1983, "Counting Rates For Representative AXAF Imaging Instruments" (Cambridge: Smithsonian) SAO-AXAF-83-017.

N87
24263
UNCLAS

16

X-RAY ASTRONOMICAL SPECTROSCOPY

Stephen S. Holt
Laboratory for High Energy Astrophysics
Goddard Space Flight Center
Greenbelt, Maryland 20771

ABSTRACT

The contributions of the Goddard group to the history of X-ray astronomy are numerous and varied. One unique role that the group has continued to play involves the pursuit of techniques for the measurement and interpretation of the X-ray spectra of cosmic sources. The latest development in this story has been the selection of the new X-ray "microcalorimeter" for the AXAF study payload. This technology is likely to revolutionize the study of cosmic X-ray spectra.

1. INTRODUCTION

X-ray astronomy is currently more than 20 years old, as measured from the first unambiguous discovery of extra-solar X-rays [Giacconi et al., 1962]. Experimental research during this first generation may be broadly classified into four main types of measurement: timing, imaging, polarimetry, and spectroscopy. This review is meant to concentrate primarily on the latter, with specific attention paid to the contributions made by the group at GSFC which was created and nurtured by Frank McDonald.

During the first decade of X-ray astronomy, timing measurements represented the most important channel of investigation. Direct measurements of periodic and Doppler-effected periodic variability signaled the very nature of the strong

X-ray emitters in the galaxy as accreting neutron stars in binary systems [cf. Schreier et al., 1972]. The value of timing measurements has not diminished, as new discoveries continue to be made via the timing channel [e.g., "QPOs", or quasi-periodic oscillations, from some bright galactic bulge sources by Van der Klis et al., 1985, using EXOSAT data], and new capabilities continue to be planned (e.g., the Japanese mission ASTRO-C and the Explorer mission XTE).

The advent of substantial X-ray imaging capability with the Einstein Observatory (HEAO-2), launched in 1978, allowed sensitivities for point source and morphological studies sufficient to make imaging an important research tool [Giacconi et al., 1979]. This trend will continue with the cooperative German/U.S. mission ROSAT, and will be considerably expanded with the advent of the major observatory AXAF.

Polarimetry is extremely difficult for X-ray astronomical sources, and only a single space-borne instrument has thus far been attempted. It achieved important confirmatory evidence for the synchrotron nature of the X-ray emission from the Crab nebula [Weisskopf et al., 1978], but that first instrument lacked the sensitivity to measure the required few percent (or better) polarization in much weaker sources necessary to allow polarization to become a more generally useful tool for X-ray astronomical research.

While all four channels offer important and complementary information, spectroscopy has steadily achieved increasing importance as X-ray astronomy has matured. This review will concentrate on the "new" X-ray spectroscopy, i.e., that which has sufficient resolving power to be concerned with measurements of discrete line features rather than just with continua.

2. INSTRUMENTAL CONSIDERATIONS

Virtually all pre-Einstein X-ray spectroscopy was performed with proportional counters, which have a resolving power limited by the nature of the atomic interactions in the counter gas (which follow the primary photoabsorption) to $R = E/\delta E_{FWHM} \sim 3E_{keV}^{1/2}$. Early attempts by the GSFC group to search for Fe emission with such counters [Holt, Boldt, and Serlemitsos, 1969] led to the development of large multi-wire proportional chambers which were

ultimately rewarded with positive detections of thermal Fe emission from supernova remnants [Serlemitsos et al., 1973] and fluorescent Fe emission from cold material surrounding the ionization source in galactic X-ray binary sources [Serlemitsos et al., 1975].

Gas scintillation counters, which shortcut the atomic interactions subsequent to primary photoionization, are limited to a resolving power which is higher by a factor of two. Such detectors have successfully been flown on the European EXOSAT and Japanese "Tenma" missions.

Even higher resolving powers in photoelectric detectors can be achieved by using semiconductors instead of gas as the detection medium. The most successful such instrument for X-ray astronomy has been the SSS (solid-state spectrometer), a cryogenically cooled silicon chip with resolving power $R \sim 6E_{keV}$ provided by the GSFC group to the Einstein Observatory (HEAO-2). Like other charge collecting proportional devices, Si(Li) can have detection efficiency approaching unity over its entire bandpass simultaneously. Dispersive devices, such as the Einstein FPCS (focal plane crystal spectrometer, with $R \sim 100$) and the OGS (objective grating spectrometer, with $R \sim 30$), have already demonstrated great analytic power with observations of a few of the brighter sources. Such instruments can be designed to have $R > 1000$, but will require very large collecting areas to be generally useful for a large number of sources because of their low effective photon detection efficiencies. Next generation FPCS and OGS systems have been conditionally selected for the initial complement of AXAF instruments.

Similarly selected is a novel new X-ray spectrometer currently under development at GSFC which is capable, in principle, of combining the high efficiency and bandwidth of the photoelectric devices with the high resolving power of the dispersive devices. The "microcalorimeter" consists of a supercooled chip in which an X-ray is photoelectrically detected and its energy is totally thermalized, with the resultant $\sim 10^{-16}$ joules measured via the rise in temperature of the chip. The system is expected to exhibit a resolving power $R > 1000$ for Fe–K lines.

The problem of translating from photons detected in a spectrometer to a source spectrum is not trivial, and it is important to recognize that the interpretation of an observation can depend upon the analysis procedure employed.

The raw data are a convolution of the actual input photon spectrum with the response function of the spectrometer, but their inversion to a derived input spectrum is not unique.

The conventional inversion is a model-dependent procedure that requires the observer to have some a priori knowledge of the actual spectral form, so that it can be characterized by a limited number of adjustable parameters. Typical fitting parameters are amplitude and shape of continuum (e.g., power law index with exponential high energy cutoff), strengths (and possibly energies) of emission lines and photoelectric absorption edges, and low energy photoabsorption by intervening cold matter. Simulated detector count spectra are computed from the assumed spectral forms, and the model parameters are varied to achieve the best fit to the actual detector counts. With this procedure, spectral features blurred by the detector response can be enhanced for display, as illustrated in Figure 1. Its hazard is that unanticipated features are forced to be represented by the assumed parameters so that, for example, an intrinsically broad line might emerge as a doublet if only sharp lines are assumed. If the lines are totally resolved, however, much of this ambiguity disappears.

3. X-RAY SPECTRAL CATEGORIZATION

The variety of X-ray sources present a variety of observational manifestations; in some cases, X-ray spectra provide unique opportunities to gain crucial insight into the nature of sources.

Since line emission ceases to be the dominant cooling mechanism of astrophysical plasmas for temperatures exceeding ~ 1 keV, X-ray spectra are primarily characterized by their continua. The simplest (at least spectrally) of these are the featureless power laws produced by the interaction of power law distributions of cosmic ray electrons with ambient magnetic fields. The Crab nebula, for example, is the archetype of such a pure synchrotron source, in which we observe the single interaction of astrophysical particles and fields.

Almost as simple, in the sense that it is also a single-parameter continuum, is the blackbody spectrum which characterizes the other extreme, i.e., the

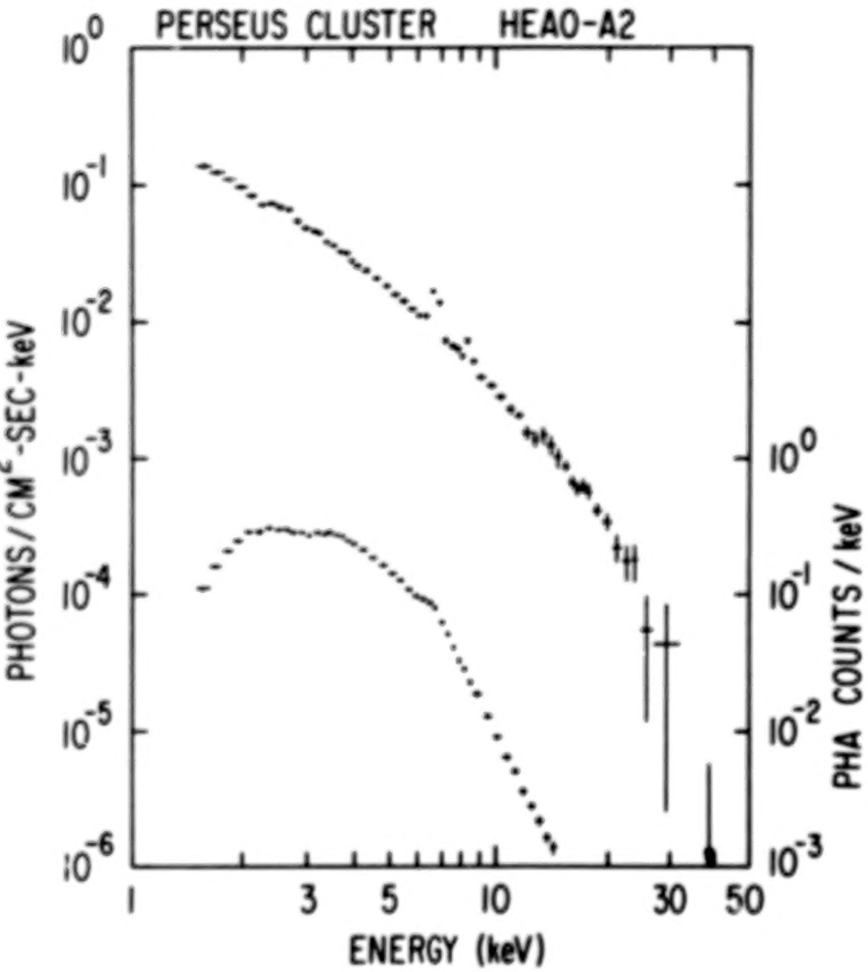


Figure 1. Data from an exposure to the Perseus cluster of galaxies taken with the Goddard HEAO-1 A2 proportional counter experiment. The lower points, for which the right-hand scale is appropriate, are the raw data in energy-equivalent pulse-height channels. The higher points, for which the left-hand scale is appropriate, represent the most probable input spectrum consistent with the assumption of an equilibrium thermal spectrum. Note, in particular, that the Kα and Kβ Fe emission lines at 6.7 keV and 7.9 keV can be enhanced for display by this model-dependent spectral inversion process.

manifestation of sufficient interactions between particles and photons that complete thermalization has occurred. X-ray "bursts" arising from nuclear burning episodes on the surfaces of neutron stars, for example, appear to exhibit these classic spectra.

The preponderance of X-ray sources typically exhibit intermediate continua in the sense that electron scattering plays a role in the formation of the observed spectra. This is true for the modified bremsstrahlung of the galactic binary systems which contain white dwarfs, neutron stars, or black holes, as well as for the active galactic nuclei which usually display a characteristic power law spectra with index $\alpha \sim 0.7$. Of particular relevance to this paper are the fluorescent line features which have been observed from both types.

Finally, the optically thin "thermal" spectra at "X-ray temperatures" of a variety of astrophysical system types provide the richest line spectra available to the next generation of instruments for X-ray spectroscopy. Optically thin plasmas offer a well studied starting point for comparison with actual X-ray source spectra. Although equilibrium and isothermality over the whole source volume may be a simpler situation than can be expected to obtain in general, less ideal sources can be modeled as a distribution of gases which exhibit thermal characteristics. The gas may be in local collisional equilibrium, so that the electrons will have local Maxwellian distributions with kinetic temperature T, while ions of charge Z may have the population fraction they would have in collisional equilibrium at a different temperature. The gas can have bulk motion, and parts of the source can be moving relative to one another. In complex sources, a nonthermal continuum may need to be taken into account, as well.

4. OPTICALLY THIN THERMAL SOURCES

In many astrophysical contexts, an X-ray-emitting plasma is sufficiently transparent that the emergent spectrum faithfully represents the microscopic processes occurring in the plasma. At temperatures $T > 10^8$ K, almost all abundant elements are fully ionized, and the X-ray emission is dominated by bremsstrahlung from hydrogen and helium. At lower temperatures,

however, trace elements retain a few atomic electrons, and their contribution to the emissivity of the plasma cannot be ignored. Because cross-sections for electron impact excitation of such ions far exceed those for bremsstrahlung or radiative recombination, line emission from $Z > 7$ constituents actually dominates the cooling from a plasma with $10^4 < T < 10^7$ K, even though these trace elements represent only $\sim 10^{-3}$ of the plasma composition by number.

A theoretical model is needed to infer physical parameters such as temperature, element composition, and density from the observed spectrum of an optically thin plasma. To construct such a model requires knowledge of the ionization state of each element, which is controlled by electron impact ionization and radiative and dielectronic recombination. If the plasma is maintained at constant temperature and density for a long time compared to the timescales for these microscopic processes, the ionization of trace elements relaxes to a stationary state that depends primarily on electron temperature and weakly on density. This stationary ionization balance assumption was first employed to interpret the solar corona, and many such "coronal models" have been calculated. An exemplary model spectrum is shown in Figure 2.

Such spectra are rich in emission lines and offer the opportunity to deduce many properties of the emitting gas when sufficient resolving power and sensitivity are available. The main emission complexes from $K\alpha$ transitions of helium-like and hydrogenic ions of abundant elements such as O, Ne, Mg, Si, S, Ar, Ca, and Fe, become distinct with resolving power $R > 10$. A temperature may be inferred from the relative strengths of the helium-like and hydrogenic lines from a given element, but its validity depends on that of the stationary ionization balance assumption. A measure of temperature that is independent of that model assumption may be obtained from the ratio of $K\beta$ to $K\alpha$ transitions of a single ion if there is sufficient sensitivity to separate the weaker $K\beta$ from $K\alpha$. With $R > 100$ a variety of more powerful temperature and density diagnostics become available. Satellite lines resulting from dielectronic recombination may be used as temperature diagnostics, and the $K\alpha$ complexes of helium-like ions can be resolved into their resonance (2^1P), intercombination (2^3P), and forbidden (2^3S) components, where the relative strengths can be used to infer electron temperature and density.

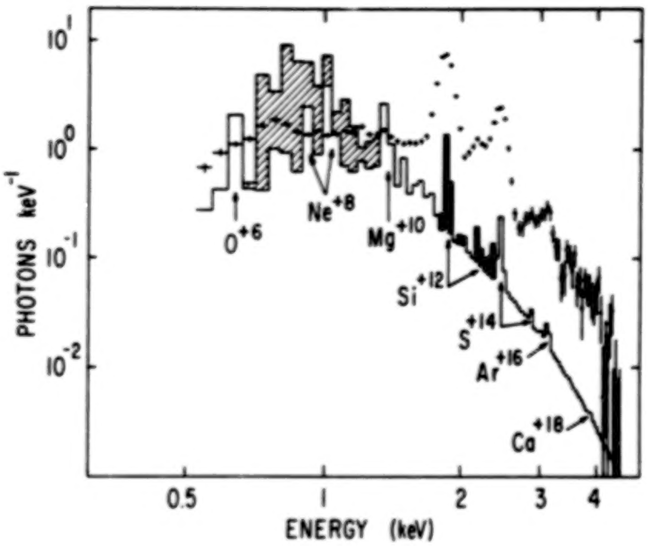


Figure 2. A comparison of the actual raw data from Tycho's SNR taken with the Goddard HEAO-2 solid-state spectrometer experiment and an idealized input spectrum. The data can be fit equally well with Sedov-blast-wave and two-temperature-equilibrium modeling, provided that the abundances of the even-Z elements are treated as free parameters. The idealized input spectrum represents the dominant $kT = 0.5$ keV component of the two-temperature fit to the data (the higher temperature component has $kT \sim 4$ keV), but with the abundances fixed at solar proportions. The shaded area represents the contribution from Fe L-blend emission at this temperature, and the blackened area represents the contribution form Si K-emission components. To facilitate comparison with the raw data (where the abscissa should properly be energy-equivalent pulse-height), the idealized input spectrum is viewed through a column density of 2×10^{21} atoms cm^{-2}. Both the data and the model are displayed in 46 eV bins, with the latter smeared to an effective resolution with FWHM ~ bin width (i.e. ~ 3 times better than the actual SSS FWHM resolution). The $K\alpha$ and $K\beta$ transitions of helium-like ions of Mg, Si, S, Ar, and Ca indicated on the model spectrum are clearly evident in the data. Since most of the K-emission arises from the lower temperature component of the two-temperature fit, it is clear that consistency with this model requires marked overabundances in the line-emitting species.

Without resolving power sufficient to determine a temperature distribution unambiguously, this distribution and the elemental abundances in the plasma must be non-uniquely determined from comparison with the data expected from a grid of models. For radiatively ionized or nonequilibrium plasmas, the model results are only as good as the reality of the model spectra. If a true equilibrium temperature distribution can be inferred, however, abundances can be determined by the comparison of lines of one element to those of other elements or to the continuum. Such abundance determinations are relatively straightforward for equilibrium plasmas, because each elemental constituent is present in just a few ionization states.

Information from less idealized situations can be gleaned with sufficient insight even when the optically thin equilibrium scenario is untenable. In the optically thin case, for example, inconsistency of the electron and ionization temperatures can be used to determine the extent to which the plasma is recombining or ionizing in a nonequilibrium situation. Colder material around or along the line of sight to the primary ionization can yield fluorescent lines, and optically thick plasmas can yield lines that are broadened by electron scattering. With sufficient resolving power, it may even be possible to measure the natural width of emission lines.

Many of the likely possibilities for X-ray spectra from astronomical objects, and the important information that the spectra might therefore reveal, can be found in reviews such as Holt and McCray [1982], from which much of the above introductory discussion has been taken.

5. REQUIRED RESOLVING POWER

A crucial consideration for any investigation is the sensitivity required to carry it out. For the X-ray spectroscopy of cosmic sources, a variety of parameters contribute to this sensitivity: the quantum efficiency of the detectors, the detector background, the detector bandpass, and the resolving power. If an "ideal" spectrometer with virtual unit efficiency, trivial background, and wide bandpass can be developed, then the only parameter which requires discussion is the resolving power.

Atomic physics prescribes the resolving power necessary to utilize various combinations of spectral lines for scientific study. For each element, we can approximate the resolving power required to separate a number of important ones. For n = (2 to 1) transitions, for example, the most important lines are those from the fluorescence of neutral material, the analog of Lyman α from hydrogen-like material, and the resonance, forbidden, and intercombination lines of helium-like material. The energy of Lyman α is, of course, Z^2 times the 10.2 eV for hydrogen. Table 1 gives the separation energies of these five lines for elements ($8 \leqslant Z \leqslant 26$) which are likely to be of interest for AXAF.

Table 1 demonstrates that while $\sim$ 1 eV may be required to completely resolve all the lines of potential interest from all elements, $\sim$ 10 eV is sufficient to separate the most important lines from oxygen, and can totally resolve them for iron.

TABLE 1: LINE SEPARATIONS

Line pairs	Approximate energy separation (eV)
Lyman α — Resonance	$10\ Z = 100\ Z_{10}$
Resonance — Intercombination	$0.32\ Z^{4/3} \sim Z$
Resonance — Forbidden	$0.77\ Z^{4/3} \sim 2\ Z$
Resonance — Neutral	$2.3\ Z^{3/2} \sim 10\ Z$

There is a similar $\sim$ 10 eV requirement commensurate with possible sources of line broadening. The natural width of a resonance line is $\sim 10^{-2}Z_{10}^{4}$ eV, so that resonance and even narrower forbidden lines have widths that are not measurable with $\sim$ 1 eV. Thermal broadening at $\sim 2(T_8 Z_{10}^{3})^{1/2}$ eV would similarly require sub-eV resolution to measure, even for the heaviest elements at the highest temperatures. Broadening associated with mass motion is easily discernible in many astrophysical contexts with 10 eV, however, as the broadening for this case is $\sim 4v_{1000}Z_{10}^{2}$eV (v_{1000} in units of 1000 km s^{-1}).

It would appear, therefore, that the few-eV resolution required for the study of all the diagnostics in Table 1 is well matched to that required for the motion-broadening in active galactic nuclei (AGN), young supernova remnants (SNR), and the strong stellar winds of early type stars. The resolving power required to measure thermal or natural broadening would be orders of magnitude better, however.

6. AN "IDEAL" SPECTROMETER

The most important attributes of a spectrometer depend upon the specific scientific objectives of a particular investigation, but there are some which are so generally useful that they can be safely assumed to be characteristic of the "ideal" spectrometer for AXAF. These attributes include energy resolution sufficient to address the most important scientific objectives for all classes of sources and sensitivity to the entire AXAF bandpass simultaneously with both near-unit efficiency and trivial background.

The X-ray calorimeter (see Figure 3) which has been selected as part of the AXAF study payload promises these attributes. It consists of an X-ray absorber of heat capacity C loosely connected to a heat sink of temperature T by a thermal link with conductivity G. When a photon of energy Q is absorbed, it is degraded into quanta (phonons) characteristic of T. The temperature of the absorber rises by an amount $\delta T = Q/C$ and then decays back to its equilibrium temperature with a time constant $\tau = C/G$. The temperature increase is detected by a thermistor attached to the absorber, and the incoming photon energy can be deduced by the magnitude of δT.

The small heat deposition is measurable to a precision which is determined by the random exchange of energy between the detector and its heat bath through the thermal link. An oversimplified "explanation" of the situation is that the effective number of phonon modes contributing to the heat capacity and to the fluctuations is $N = C/k$, each with effective mean energy kT. If the typical quantum occupation number in each mode is unity, the rms fluctuation in that number is also unity. Therefore, the total energy fluctuation in the system is $\delta E \sim kT(N^{1/2}) \sim (kT^2C)^{1/2}$.

X-RAY CALORIMETER CONCEPT

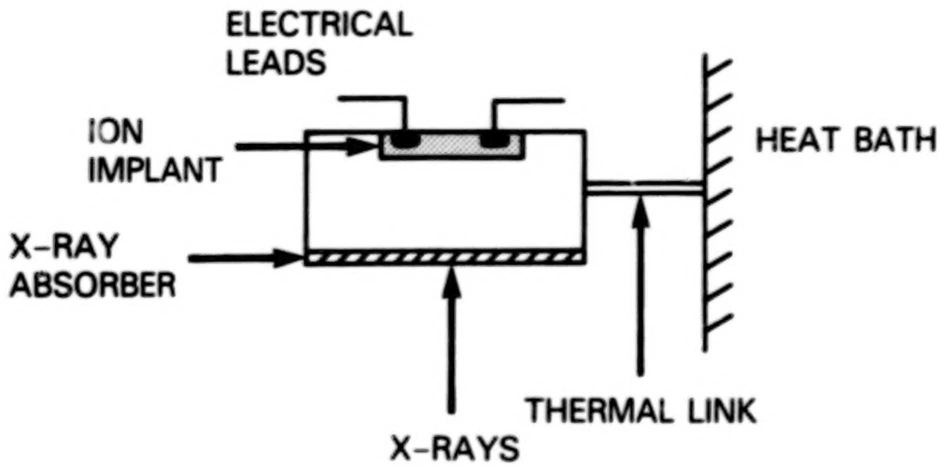


Figure 3. Schematic of an X-ray calorimeter. The essential elements are an X-ray absorber, a temperature transducer, and a thermal link to a heat bath which is loose enough to ensure that the rise in temperature accurately reflects the deposited energy (i.e., that the timescale for total thermalization of the X-ray energy is short compared with the characteristic time constant of the thermal link).

Because the number of phonons involved in the final steps of degradation of phonon energy to heat is large, the statistical fluctuations which fundamentally limit the resolution of charge collecting detectors (such as Si(Li) or proportional counters) do not significantly affect the temperature increase produced by a single photon; this means that the limiting FWHM resolution of the device will be uniform over the whole bandpass.

When all necessary noise contributions (such as load resistor Johnson noise and realistic filter techniques) are taken into account, the limiting resolution of a practical detector is δE (FWHM) $\sim 4(kT^2C)^{1/2}$ [Moseley, Mather, and McCammon, 1984], which can be as low as 1 eV for a detector with total heat capacity $< 10^{-14}$ J/K operating at 0.1 K. Such detectors can be designed

for AXAF, but it remains to demonstrate that the theoretically predicted performance can actually be obtained.

7. RECENT PROGRESS IN DETECTOR DEVELOPMENT

There are two general types of noise source which can potentially prevent the achievement of the theoretical performance discussed above. The first might be called "conversion" noise, as it arises from the imperfect conversion of X-ray energy to phonons, e.g., energy which goes into electric charge or which is trapped in states with long lifetimes. The second might be called "readout" noise, as it arises from imperfect transduction of the temperature increase into interpretation as an energy deposition.

The latter potential noise source may be ultimately limiting in attempting to attain 1 eV, but it has already been experimentally demonstrated to be < 10 eV via the observation of fluctuations in the sampling of the baseline of a test detector with heat capacity higher than the value required for 1 eV [Moseley et al., 1985]. Since we could not simultaneously demonstrate similarly low conversion noise, our efforts during the past year have been aimed at understanding and reducing this component. We have recently succeeded in achieving that level of conversion noise via the comparison of X-ray resolution with conversion-independent baseline fluctuations.

Figure 4 demonstrates the performance of a nonoptimized composite detector with baseline fluctuations of ~ 30 eV FWHM. An overnight run was performed with an Fe^{55} source, yielding a total additional noise contribution (including conversion noise) of < 10 eV in the K-capture lines and in fluorescence from an Mg target. It is interesting to note that the $K\alpha$ line is better modeled with its two components ($K\alpha1$ and $K\alpha2$) than with a single line, and that their separation of ~ 10 eV is at about the current level at which we can demonstrate the contribution from either conversion or readout sources (although the latter contribution in this particular detector is three times larger).

There still remain subtle problems with the production of 1 eV detectors, but we have now demonstrated that 10 eV detectors can be made. Our goal is

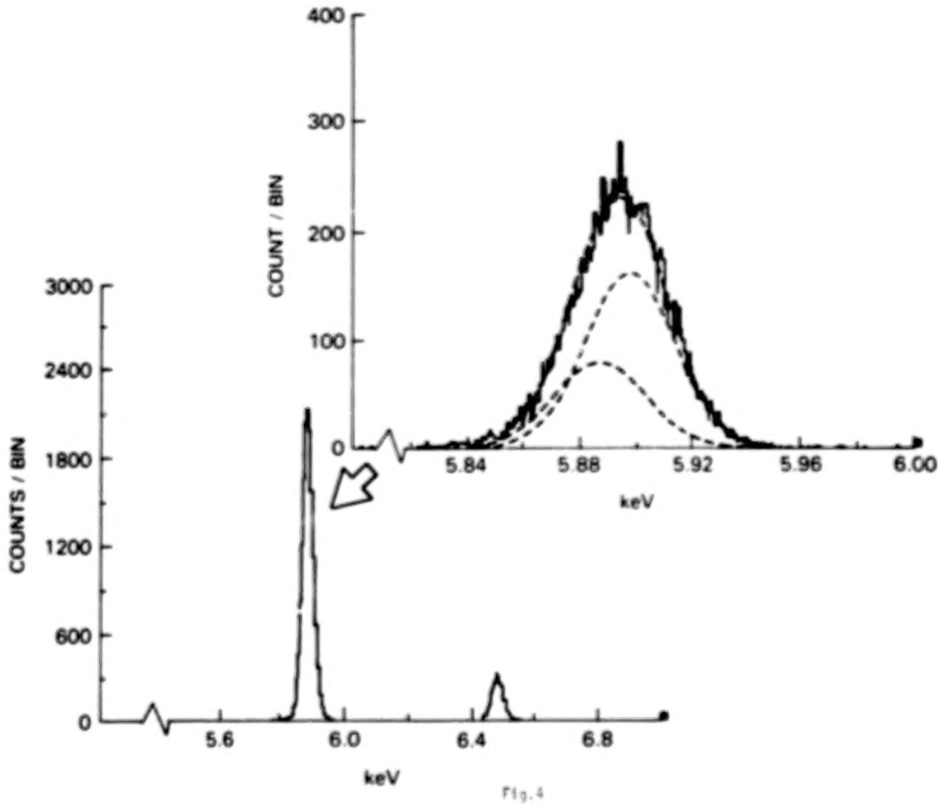


Figure 4. Raw data from an Fe^{55} X-ray source taken with a composite calorimeter consisting of a HgCdTe X-ray absorber on a Si substrate with an ion-implanted thermistor. The Kβ line is ~ 600 eV higher in energy than Kα, and an order of magnitude less prominent. The Kα line has two major components separated by ~ 10 eV, with the higher energy component approximately twice as prominent. These two Kα components are required for an acceptable fit to the data, but cannot be resolved by the current ~ 30 eV detector. The AXAF detector should be capable of completely resolving them.

to produce detectors for AXAF for which Kα1 and Kα2 from Fe^{55} will be completely resolved. Note that the spatial resolution of the AXAF telescope (< 1 arc-sec ~ 0.05 mm) is required in order to reach this goal, as the necessary value of C cannot be attained without a sub-mm sized detector.

8. SUMMARY

The AXAF microcalorimeter provides a totally new capability for X-ray astronomical spectroscopy. The combination of resolving power of order 10^3 with quantum efficiency of order unity (with virtually insignificant detector background) simultaneously over the entire AXAF bandpass allows for the application of X-ray spectroscopy to much more ambitious scientific studies than previously possible.

The X-ray astronomy group at Goddard has made numerous contributions to astrophysics, and no subject represents its experimental motivation more than does X-ray spectroscopy. It is particularly fitting for the purposes of this volume that the microcalorimeter, the latest manifestation of this experimental research program, has arisen out of a true collaboration between the Goddard X-ray and infrared groups, both of which owe their existence to the foresight of Frank McDonald.

REFERENCES

Giacconi, R., Gursky, H., Paolini, F., and Rossi, B., 1962, *Phys. Rev. Letters.*, **9**, 439.

Giacconi, R. et al., 1979, *Astrophys. J.*, **230**, 540.

Holt, S. S., Boldt, E. A., and Serlemitsos, P. J., 1969, *Astrophys. J.*, **153**, L137.

Holt, S. S., and McCray, R., 1982, *Ann. Rev. Astron. Ap.*, **20**, 323.

Moseley, S. H., Kelley, R. L., Mather, J., Mushotzky, R. F., Szymkowiak, A. E., and McCammon, D., 1985, *IEEE Trans. Nucl. Sci.*, **NS-32**, 134.

Moseley, S. H., Mather, J., and McCammon, D., 1984, *J. Appl. Phys.*, **56**, 1263.

Schreier, E., Levinson, R., Gursky, H., Kellogg, E., Tananbaum, H., and Giacconi, R., 1972, *Astrophys. J.*, **172**, L79.

Serlemitsos, P. J., Boldt, E. A., Holt, S. S., Ramaty, R., and Brisken, A. F., 1973, *Astrophys. J.*, **184**, L1.

Serlemitsos, P. J., Boldt, E. A., Holt, S. S., Rothschild, R. E., Saba, J. L. R., 1975, *Astrophys. J.*, **201**, L9.

Van der Klis, M., et al., 1985, *Nature*, **316**, 225.

Weisskopf, M. C., Silver, E. H., Kestenbaum, H. L., Long, K. S., and Novick, R., 1978, *Astrophys. J.*, **220**, L117.

N87
24264
UNCLAS

N87-24264

17

INFRARED ASTRONOMY

Michael G. Hauser
Laboratory for Extraterrestrial Physics
Goddard Space Flight Center
Greenbelt, Maryland 20771

1. INFRARED ASTRONOMY

One of the lesser known contributions of Frank McDonald to space science is his role in extending high energy astrophysics to the sub-eV photon energy range—in putting infrared astronomy into orbit. This seemingly paradoxical extension was, in fact, both reasonable and timely. Reasonable, because so many infrared photons emanate from highly energetic sources such as star-forming clouds, active galaxies and the Big Bang itself. Though each photon carries modest energy, one must measure the infrared brightness to be able to understand overall source energetics. Timely, because the technical ingredients needed to make infrared space astronomy a reality, including propulsion, attitude control, cooling, and sensors, were either in hand or at least within reach.

In the spring of 1974 I was working at Caltech on new detector technology for far-infrared astronomy. I had the good fortune of having lunch one day with Len Fisk, then a member of McDonald's High Energy Astrophysics Division at Goddard. Len mentioned that Frank was thinking that it was time to start a serious effort leading toward space astronomy in the infrared, and

suggested that I contact Frank if that would interest me. I arrived at Goddard late in the summer to take up the challenge, buoyed substantially by my confidence in Frank's support and judgment, already well-confirmed in numerous other disciplines.

In addition to enthusiasm and moral support, Frank bestowed upon his fledgling effort several indispensible ingredients: talented, highly motivated people, in this case three individuals from other parts of his Division (Bob Silverberg, just completing his Ph.D. research on cosmic rays; Dave Walser, an experienced technician about to complete his electrical engineering degree; and Tony Flanick, the senior technician in the low energy cosmic ray group); funds to begin equipping a laboratory and developing hardware (I never asked him whose account he spirited this from); and space for our laboratory (those who have not worked in Building 2 cannot appreciate the magnanimity of this contribution). In his quiet, unobtrusive way Frank offered much valuable advice, ranging from whom to contact at Goddard or Headquarters to make things happen (he immediately put us in contact with Pat Thaddeus at the Goddard Institute for Space Studies (GISS), with whom we have had fruitful collaboration to this day), to how to make recruiting decisions (put talent and quality of work above specialized experience). Finally, of course, he provided a stimulating and fertile work environment, characterized by the intellectual fervor of the many research groups with overlapping scientific interests, solid engineering and software support, and unflagging management encouragement with minimal bureaucratic distraction.

That Frank's judgment was impeccable as to the proper timing for pursuit of infrared astronomy from space is borne out by the subsequent record. By the time I arrived at Goddard, NASA had announced an opportunity to propose Explorer-class missions or mission concept studies, which turned out to be the only such opportunity in the next dozen years. At that time, the infrared sky was little known at wavelengths longer than 20 microns, and was unobserved on spatial scales larger than tens of arc-minutes at any wavelength because of the bright atmospheric emission. At the long wavelength end of the infrared spectrum, there was intense cosmological interest in measuring the shape of the cosmic background radiation spectrum, then firmly established

only on the Rayleigh-Jeans side of the presumed 3 K spectrum, and in determining the anisotropy of that radiation.

With Frank's support and guidance we generated two proposals within a few months: one to conduct a sensitive sky survey at 50 and 100 microns, and the second, initiated by Pat Thaddeus and John Mather, then a young post-doc at GISS, to build a Cosmic Background Radiation Satellite to make definitive measurements of the microwave background and search for a primeval infrared background component. The latter included as collaborators David Wilkinson from Princeton and Rainer Weiss from MIT. In retrospect, these proposals were very much on target scientifically, but were programmatically remarkably naive (for example, we seriously discussed building the entire instrument complement for the cosmic background mission, including the dewar, for five million dollars!). Fortunately, the estimates became more credible, the scientific promise prevailed in the evaluation process, and the first proposal led to our participation in the Infrared Astronomical Satellite (IRAS) mission, while the second, though via a more protracted path, led to the Cosmic Background Explorer (COBE). We were suddenly caught up in the process of opening a new window on the universe, and in probing its most ancient radiative relics!

The road to completion of space programs is a lengthy one, and these two in particular have redefined the meaning of Explorer-class missions. With the wisdom of experience, and perhaps a bit of prescience, Frank from the outset encouraged us also to pursue programs which would provide both short-term scientific return and opportunities to test new technology germane to the space business. Our first such endeavor, also begun in the fall of 1974, was development of a balloon-borne telescope and submillimeter photometer for investigation of the large-scale luminosity and interstellar mass distributions in the galactic plane. With pointing control and maneuvering requirements substantially tougher than those of previous balloon payloads developed in Frank's lab, we presented a significant challenge to his engineering staff, as he no doubt wished. Their able contributions, along with those of several other engineering groups at Goddard, resulted in a payload which delivered high quality maps of the inner galaxy in each of its three flights. These data (Figure 1) show the high infrared luminosity of the inner galaxy, and, together with the data from the CO survey of the same region by Thaddeus and his colleagues,

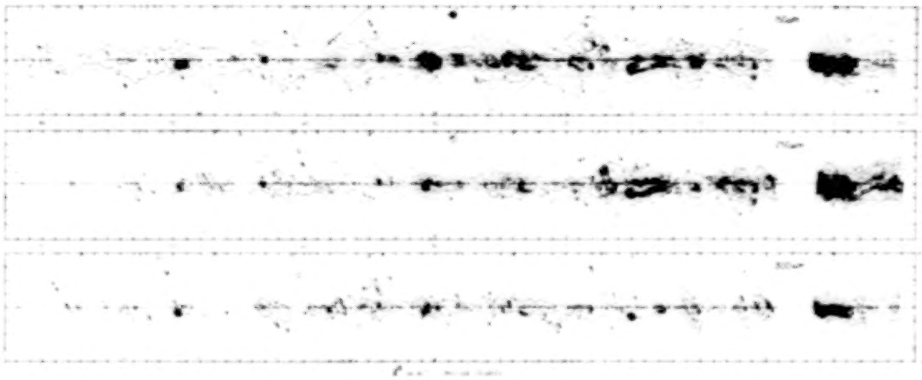


Figure 1. Contour maps of galactic plane emission at 150, 250, and 300 microns obtained with the 1.2-m diameter balloon-borne telescope developed by the Infrared Astronomy Group in the High Energy Astrophysics Division. Extensive diffuse emission is clearly evident, punctuated by discrete sources which are typically associated with prominent HII regions. Though it is generally agreed that the diffuse emission is thermal emission from dust, the location of the dust, whether predominantly associated with molecular, ionized, or atomic gas remains a question of study. [Figure originally presented by Hauser et al., 1984, Astrophys. J., **285**, *74.]*

have allowed systematic study of the distribution and properties of major star-forming regions in the galaxy [Hauser et al., 1984; Myers et al., 1986].

Subsequent to these flights, the IRAS mission did, of course, initiate the era of space infrared astronomy in spectacular fashion [Neugebauer et al., 1984; Hauser, 1986]. We have, for the first time, complete imagery of the sky in the infrared (Figure 2), including an initial determination of the heretofore elusive absolute brightness of the sky, and catalogs of hundreds of thousands of discrete sources, including tens of thousands of extragalactic objects. New insights and discoveries have emerged in wide ranging astrophysical domains, from concentrations of dust in the solar system offering clues to the origin of the interplanetary cloud, to primordial solid material orbiting nearby stars suggesting early stages of planetary system formation, and to galaxies with near-quasar luminosity emitting more than 99% of their energy in the infrared.

Figure 2. The distribution of 100 micron emission over the full sky as measured by IRAS. This Aitoff projection is in galactic coordinates, with the galactic center at the center of the projection. The data have been filtered to suppress the very large-scale emission, such as that from interplanetary dust. In addition to the prominent emission from the galactic plane, here seen all around the sky, one sees emission from the nearby molecular cloud regions such as Ophiuchus, Orion, and Taurus, large loop structures seen also in HI maps of the galaxy, patchy clouds of emission dubbed 'infrared cirrus' by the IRAS team, and some prominent external galaxies such as the Magellanic Clouds. Space-borne cryogenic telescopes are essential to obtaining such a global view in the infrared. (Figure prepared from IRAS data by J. Good, JPL.)

The scientific harvest from IRAS will continue for many years as the astronomical community scrutinizes and ponders the data.

When NASA decided to begin serious study of the COBE mission, Frank McDonald immediately rose to the occasion by urging that we try to hire John Mather, who was then about to take a position at Bell Labs. That was probably the second best decision in the program (the first being to do it, of course!), because John could not resist the temptation to pursue the secrets embedded in the cosmic background spectrum, and has been a mainstay of the program throughout its many phases. The program not only promises uniquely exciting science, it has also, with the decision to build the instruments and spacecraft in-house, turned out to be a major contributor to the maintenance and development of Goddard's engineering capabilities. The COBE hardware is now in an advanced stage of development, and we eagerly look forward to the characterization of the cosmos which this mission will provide.

Frank McDonald's tenure as the official champion of the infrared astronomy group which he formed was relatively brief. About three years after the group was formed, we were reorganized into the Extraterrestrial Physics Division under the able guidance of Norman Ness. Frank himself subsequently left Goddard to take the position of Chief Scientist at NASA Headquarters. One might suppose that under these circumstances his contributions to our discipline have diminished. I have quite a different view. Both through intangibles, such as the quality and spirit which remain in the group which he created, and in the very tangible support and encouragement he provides for the ongoing programs established under his stewardship and for realization of the next big step in this field, the Space Infrared Telescope Facility, he continues to be a major asset to our endeavors. Frank McDonald's name may never appear on any of the research papers in infrared astronomy; nevertheless, he is leaving his mark on the field no less surely than on the many others which he has inspired and led.

REFERENCES

Hauser, M. G. et al., 1984, *Astrophys. J.*, **285**, 74.

Hauser, M. G., 1986, "The Cosmos According to IRAS", in *Yearbook of Science and the Future*, (Encyclopedia Britannica, Inc.), p. 28.

Myers, P. C. et al., 1986, *Astrophys. J.*, **301**, 398.

Neugebauer, G. et al., 1984, *Science*, **224**, 14.

IV

AFTERWORD

PRECEDING PAGE BLANK NOT FILMED

WELCOMING TALK

Noel W. Hinners
Director, Goddard Space Flight Center
Greenbelt, Maryland 20771

It is really appropriate and also a pleasure to welcome you here today, in order to give particular thanks to Frank McDonald—or at least his mother and father—for participating in a cosmic event that took place about 60 years ago. Following his birth, there was a gap in Frank's life history that I haven't figured out yet, but he got his bachelor's in 1948 at Duke University and went on to that hot bed of radicals in Minnesota, to get his master's and Ph.D. Before we continue, it is worth refreshing your memory as to what can influence a person's background; those of you who are familiar with the origins of the space program know what sort of characters came out of Minnesota. Frank's credentials include an Atomic Energy Commission fellowship in 1951-53, and a research associateship in another interesting place, the University of Iowa. Those were the right places to be in those days, and he followed up with the right move: he came to the Goddard Space Flight Center in 1959. After he had been made an assistant professor at Iowa, he formed his group here and headed the Energetic Particles Branch for the next 11 years. He did a tremendous job initiating work on many of those early Explorer missions. Frank was, in fact, principal investigator on 15 of them. He did so well as a branch head that he was made chief of the Laboratory for High Energy Astrophysics in 1970, doing that job until 1982. During those years, Frank was project scientist at Goddard in over 11 missions. He set the scale to what a project scientist should be, which is foremost a productive scientist, one that is respected by his peers and colleagues. For that example, Frank, we thank you. Frank finished up these projects as study project scientist for the Gamma Ray Observatory. He did well in getting GRO started, so we'll be watching that one soon, hoping that with a little luck it will turn out to be a really impressive mission. Frank has many awards, and has participated in a large number of professional societies—APS, AGU, AAS—again, the mark of a very productive man, interacting with the external community. Frank has also served on all kinds of committees, commissions, and working groups, trying to plan

for the future of astrophysics, creating those opportunities for making the great new discoveries. All these are documented facts, but, if you look back from this perspective in a federal laboratory, you will see what is unique in Frank's contribution. He has created a laboratory for astrophysics through his own initiative, scientific qualities, and productivity. A most important part of that over the years has been the nurturing of graduate students and young scientists. He has created leaders who took over Frank's leadership role when Frank moved on. Frank, to sum it all up, you have had a truly remarkable and productive scientific career. We are proud of what you have brought to Goddard and we are proud to honor you today on 60 years of productivity. Thank you.

INTRODUCTORY REMARKS

John A. Simpson
Enrico Fermi Institute
University of Chicago
Chicago, Ilinois 60637

We are here to celebrate Frank's contributions to science and to recognize his services to the scientific community that have done so much to benefit space science over the years. Many of you, especially the younger members, may not realize that when NASA got started and the Goddard Space Flight Center came into being in the late 1950's, it was an era when great opportunities in space flight existed but the program planning was in somewhat of a mixture. So it was with particular excitement that Frank spearheaded many efforts at Goddard. Many of us in the universities participated in the implementation of whole series of new spacecraft, doing work in 'fields and particles' as well as in other areas that ultimately were to branch out into new disciplines in astronomy and space research. The most outstanding example that I would choose is the series of Interplanetary Monitoring Platforms, the IMP series. I remember very well when Frank came to me, and I'm sure he talked to some of you, discussing the possible development of new instrumentation at the beginning of this program about 1961. The first IMP, launched in November 1963, led to many significant discoveries, some of which you will hear about today, covering topics from the interplanetary structure and the Earth's shock front to solar cosmic ray particle behavior. But that was only the beginning. Although, from the point of policy, it was never an explicit part of NASA's thrust that a Center would take the lead in space research, it was very important that Goddard would join hands with the universities and the various national laboratories to become so involved. For example, in the case of the IMP missions, that series went up to number 8, and IMP-8 is still operational and doing very well. I emphasize this because, although almost all of you must know of Frank's contributions to the areas that you are interested in, you may not be aware of his historic deep and firm commitment to the association with the participating scientific community outside of NASA. So I congratulate both Frank McDonald and the Goddard Space Flight Center for having the wisdom to draw science from wherever there are contributions to be made.

CONCLUDING REMARKS

Phyllis Freier
Department of Physics
University of Minnesota
Minneapolis, Minnesota 55455

This book will be the written record of the very happy day when Frank McDonald's friends and colleagues gathered to honor him for his contribution to Space Science. It was held at Goddard on April 23, 1985, 32 days before Frank officially reached the 60th anniversary of his birth. It was very appropriate to hold the celebration at the Goddard Space Flight Center, where Frank was a physicist and head of the High Energy Astrophysics Division from 1959 until he left in 1983 to become Chief Scientist at NASA Headquarters. So we gathered in the halls of Goddard to hear Frank's colleagues and friends talk about diverse areas of astrophysics, all of which Frank McDonald had participated in, either as a scientist or as a science administrator. You can judge for yourself the impressive scope of the work he has contributed to by reading this book or even by scanning the Contents.

It was equally appropriate to honor Frank and his style of doing business by holding the evening banquet at the Goddard Recreation Center. I had never been in that building; when I saw it, I realized that no one could accuse Goddard of using taxpayers' money for plush or fancy guest facilities. We sat around picnic tables, and stood in line around the barbecue pits to collect our food, buffet style. Everyone talked happily and loudly (you had to: no acoustic tiles there!) and had a most festive time. Frank's three children and two granddaughters were there, along with colleagues from all over the world, as well as nearly the entire high energy astrophysics staffs from Goddard and the University of Maryland. It was indeed a most festive occasion to celebrate science and friendship, and to demonstrate our respect and love for a person who has never placed much emphasis on fancy frills but who holds in greatest regard the qualities of friendship, honesty, and integrity, along with good science. So it was an occasion we all celebrated, remembering the personal

times we had shared with Frank McDonald, as earlier during the day we had heard the talks recalling the science we had shared with him.

Of all the people at the celebration, Ed Ney and I probably have had the longest associations with Frank, the physicist, dating back to when he entered graduate school. In the fall of 1948, Frank came to the University of Minnesota, fresh from Duke University, to study in the Physics Department. We had just discovered heavy nuclei in the primary cosmic rays in the previous April, so it was an exciting time at Minnesota with lots to do. I was a graduate student in charge of the nuclear emulsions and John Naugle was a senior undergraduate, working for me. At that time, three young assistant professors had been hired about a year earlier. They developed a lightweight cloud chamber to fly on balloons, to take advantage of General Mills' expertise in manufacturing high altitude balloons and Otto Winzen's expertise in flying them. Frank McDonald picked one of the three, Ed Ney, as his thesis advisor shortly after his arrival. I always like to tell the story of how I had picked out Ed Lofgren as my advisor, who was the oldest and who seemed, in my opinion, to be the most stable of the three. When Frank came, Ed Lofgren had just left to go back to his first love, California, to help build the Bevatron, the accelerator that would produce the first antiprotons. So after Ed left, I chose Frank Oppenheimer, presumably the second most stable. Not too much later, he had to resign from the university under tragic circumstances, as we all know, and in fact had to testify on the very day I had scheduled for my prelims. So Frank and I wound up with the same advisor—and, completely contrary to my expectations, Ed Ney, now the Regents Professor of Physics at the University of Minnesota, turned out to be the most durable of the trio.

Frank had come from the deep South with a really thick accent. In those days, professors often served liquor, but we graduate students usually limited ourselves to beer. Frank pronounced it with two syllables: 'beeaah'. It was not until much later that Frank learned to say "beer" properly. Frank's wife, Ginny, was a speech therapist, and whenever Frank wanted a beer, Ginny refused to give him one until he pronounced it correctly. I think Frank can say "beer" now. The people in the physics department in those days could be characterized as "straight-lacers" and "hell-raisers". Frank had no trouble deciding which group he was going to join. Frank had a car and that, along with his natural charm, made him very popular with both graduate

students and secretaries. It wasn't much of a car by the standards of today, or even for those days, as Frank realized. Once during the winter, Frank came to the physics department boasting because at a treacherously slippery intersection some big, fancy car had tried to take the right of way away from him. Frank, realizing that one more dent in a fender on his car would hardly be noticeable, had bluffed out the big car. As it turned out, it was the car of a former governor from one of the leading families in the state. Frank had no more reverence for people in high places then than he does now, and he took just as much delight in challenging authority then as now.

When he first came to Minnesota, Frank had worked in the emulsion lab with John Naugle and me for a few months, learning in this very visual way about cosmic radiation. The emulsions in those days were Ilford C-2 (the kind in which the pion was first detected) and although the threshold was not supposed to be as low as 4 I_{min}, we had indeed seen long straight tracks that we knew must be α-particles, so that one could measure their intensity in the cosmic radiation using nuclear emulsions. But at that time I thought α-particles were too near the threshold, and anyhow we had so much yet to learn about the much more easily detected heavy nuclei. Later when G-5 emulsions became available with their threshold below I_{min}, we indeed did use them to measure the energy spectrum of α-particles. But Frank had accepted the challenge of building a detector for cosmic ray helium nuclei, and that is what he did for his thesis problem.

Frank's thesis was titled "A Cloud Chamber—Scintillation Counter Determination of the Flux of Primary Cosmic Ray Alpha-Particles". His complete detector including a cloud chamber and scintillators weighed only 162 pounds. Frank had four flights with this detector and measured the primary alpha flux over Texas and Minnesota. From Texas he flew on February 1, 1953, recovered the load ten minutes after it touched down, and reflew it three days later. However, this time he was not so lucky. In some way the parachute had snagged and the load had free fallen from 30,000 feet. It was recovered 17 days later on a San Angelo sheep ranch, completely demolished. However, the camera and film were okay and Frank got data from this flight as well as the first one.

Undaunted, Frank rebuilt the detector and on September 16, 1953, just seven months later, he flew successfully to 18 g/cm^2 over Minneapolis. That was

standard balloon performance for those Sky Hook balloons—90,000 feet. Frank made a second flight on October 5, this time with no bad luck. He had sufficiently good data to determine the integral α-particle flux at two different rigidity cutoffs. Ed Ney still says that Frank got some of the best cloud chamber pictures that anyone acquired in a balloon experiment.

After he earned his Ph.D., Frank went to the University of Iowa to join James Van Allen. In 1957, Frank was in charge of the Iowa contingent to a Guam expedition where they first fired "racoons"—rockets launched from balloons. Frank and Kinsey Anderson, being senior members of the expedition, got to spend evenings in the Officers' Club while students like Larry Cahill and Bill Webber had to stay up late getting detectors ready to fly. Larry Cahill remembers when Frank almost got himself launched as he was jumping back and forth over the load line, enjoying his supervisory role, and the balloon was suddenly released, knocking Frank down, but—thankfully for us all—not launching him.

After Frank's decade spent in the Midwest, he went to NASA in Washington, D.C. in 1959. At this birthday celebration, I have observed the respect and genuine affection his colleagues have for him. In spite of the fact that he has taken strong and sometimes controversial stands on NASA policy matters, he seems to have made very few enemies and many, many friends. This book is a tribute from his friends to Frank McDonald—distinguished, outstanding, and widely recognized for his scientific achievements, and respected and loved by us all.

AN IRREVERENT HISTORY OF SPACE SCIENCE

John E. Naugle
Former NASA Chief Scientist
7211 Rollingwood Drive
Chevy Chase, Maryland 20815

It is a pleasure to honor my old friend and colleague, Dr. Frank B. McDonald. It is 37 years since Frank first appeared on my horizon. It was the fall of 1948, and we were taking a course in analytical dynamics from Ed Hill at the University of Minnesota. Frank was thin in those days and had a little hair on the top of his head. He wrapped those long legs around his chair to help his concentration. I spent three hours a week alternating between attention to Hill's abstruse lectures and worrying as to whether Frank would permanently become entwined with his chair and we would have to break a leg, either his or the chair's, to get him unlocked.

In those days at Minnesota, the Physics Department divided itself into the "scientists" and the "hell raisers". Frank, the good old Southern boy who had been "raised right", gravitated into the hell raisers where he began a kind of belated teenage revolt which continues to this day. As a hell raiser, you were entitled to do what the name implies, but you were also expected to get the best grades, do the best physics, and appear for work at the appointed time ready to perform perfectly no matter how late you had been out the previous night. Now there were mornings when Frank dragged a little and there were mornings when I wouldn't have given two cents for his chances of reaching sixty, even less that I would, and absolutely nothing for the chance of our both reaching sixty.

But he did and I did and here we are thirty-seven years later at the end of a long day, with everybody eyeing their watches to see how long it is till dinner time and I am supposed to say something nice about Frank, something significant about the history of space science and something amusing enough

to keep you from wandering off. Along the way I will commit heresy by saying something nice about two Government institutions, ONR and the old NACA.

I choose to insert "irreverent" in the title to warn real historians to beware. Irreverent means "deficient in veneration or respect" and in this case applies more to history than to space science. This gives me license to shade the facts a little here and there to make the story come out better. After having read over what I finally wrote, a better title might be "The Origins of the NASA Space Science Program" or better still "A Clash of Two Cultures".

I am going back to 1957 to look at the backgrounds of three kinds of people who came together in the late 1950's and early 1960's—some as NASA scientists and administrators and some as academic space scientists—who, by the policies they set, the decisions they made, the actions they took and the work they did, created the space science program of the past quarter century.

The first group I will discuss is the cosmic ray physicists who lived and worked primarily in the Midwest. I discuss them first because our hero, Frank, came from that group. It is the group I came from and know the best.

We used balloons to carry cloud chambers, Geiger counters, and nuclear emulsions as high as possible for as long as possible to catch as many cosmic rays as possible. Cosmic rays were a natural phenomenon which we wanted to understand. We wanted to measure their charge, mass, flux, energy spectrum, and the fluctuations of those qualities with time and location. We wanted to know their origin and the source of their enormous energies. Until the Bevatron at Berkeley came on line, we also used them as a source of high energy particles to generate mesons. We generally recorded our data on film, recovering both our instruments and our data on parachutes at the end of the flight. We published in the *Physical Review*, the *Journal of Geophysical Research*, the *Journal of the Philosophical Society* and *Il Nuovo Cimento*, the last affectionately known as "New Cement". We got up from Thanksgiving dinner and took the train to Chicago for the Physical Society Meeting. Sometime in January or February when the snow was deep and before spring plowing began, we went off to the Midwest Cosmic Ray Conference—a sort

of movable feast that traveled from Minneapolis to Chicago to Iowa City—where we presented informal papers, quarreled a lot about our results and their significance, ate a little and drank a lot. In the spring when the azaleas bloomed, we came to Washington to the old Sheraton Park Hotel and the Bureau of Standards, for the Annual Meeting of the Physical Society. Periodically, we went off on safaris to exotic places such as Saskatoon, Canada, Peyote, Texas, and Sioux Falls, South Dakota, to fly balloons and kill a few cosmic rays.

Of all the devices we used to catch the elusive cosmic ray, the cloud chamber was the most difficult to make work, and also the most unreliable and least productive. Frank, the rebellious Southern baptist, iconoclast, and tilter at windmills, naturally chose the cloud chamber as his cosmic ray killer.

I can still see a long gangly Frank, sitting on the floor of the cosmic ray laboratory in the basement of the physics building bent over a wire-draped recalcitrant cloud chamber with his long thin fingers entwined in the wire. It seemed that he spent years in that position but I suppose it was only six months or so. Ultimately, he got it together, and it flew and worked.

What those of us who sat around and worried over Frank's sanity and future didn't realize at the time was what an excellent training the building and operating of a cloud chamber was for a future space scientist. You had to think through your experiment from beginning to end; you had to figure out how to make it work while hanging all by itself on the end of a balloon. You had to have the fortitude to start all over again when the balloon or the parachute failed and you lost your cloud chamber and got no cosmic rays. Spending three or four years getting a Pioneer 10 instrument ready for launch on a paper-thin Atlas-Centaur filled with liquid hydrogen and oxygen, expecting it to survive a flight through Jupiter's radiation belts and sail on out of the solar system was a natural extension of this early cloud chamber idiocy.

Such was the life of a cosmic ray physicist in the Midwest prior to Sputnik. We were supported and watched over by an enlightened and benevolent Office of Naval Research which expected us to do good original physics, publish our results in reputable, refereed journals, and go on to become professors, or government or industrial scientists. We graduate students didn't worry a

lot about budgets and congressional actions, although I expect our mentors, Ed Ney, John Simpson, and Jim Van Allen did. We did not think a lot about science and public policy or government support of science; however, when we later came to NASA and began to establish policies and conduct space science we never really thought of using any policies or procedures other than those we had grown up with under ONR. Frank, Kinsey Anderson, Phyllis Freier, Bill Webber, and I were among those who grew up in that ONR culture.

You youngsters living in this modern world of carefully apportioned funds and equally carefully delineated agency roles might ask, "Why in the world was the Navy supporting balloon flights in the Midwest, which is just about as far from an ocean as you can get in these United States?" I quote from Dr. Harvey Brooks, reviewing 20 years of ONR history in 1966:

"The areas of research funded by an agency may be divided into three general classes:

a) Fields of science in which the mission orientation admits of no clear limits to agency interest, and requirements differ in both kind and volume from those of any other component of the nation's technological community.

b) Fields of science which are of vital importance to the agency mission, but whose importance is shared almost equally with other agencies of the government.

c) Fields which at present show no obvious promise as sources of concepts or research results for near-term agency exploitation but which, in the mainstream of imaginatively advancing science, can produce results of potentially significant repercussions. These areas—such as pure mathematics or elementary particle physics—may have significance either through discoveries which arise directly in the science or, more commonly, the chain of evolution of scientific ideas which links the fields of greatest scientific interest to those of extensive value."[1]

---

[1] *Research in the National Purpose*: Edited by F. Jochim Weyl, ONR; Published by ONR, 1966

ONR had a very enlightened policy regarding the importance of basic research to the national purpose. Clearly, the support of the Midwest cosmic ray physicists and their balloons was justified under class c fields of research.

The second group of people who came together to form the space science program also operated under the ONR umbrella but in fields more in the foregoing classes a and b. These were the rocketeers who operated primarily out of the Naval Research Laboratory, although Nelson Spencer had a group at the University of Michigan and Bill Stroud a group at the Army Signal Corps Laboratory. This group also studied natural phenomena, the aurora, the ionosphere, the Sun and stars, and the upper atmosphere. They used a different form of transportation, the rocket.

If you want to use a rocket for transportation you must design your instruments to fit in a very small volume and build them so they can be put on top of a rocket and withstand a controlled explosion to propel them into space. Rocketeers like a little more sound and fury in their work than those of us who used balloons. The rocketeers were willing to trade time for altitude, really they were much more than willing. They had to get above the atmosphere in order to do their work. They used telemetry to get their data back and they were the first space scientists.

People who did research using balloons or rockets really appreciated their transportation, both for its natural beauty and for its results, particularly when it worked, and there were results.

A balloon launch on a frosty Minnesota morning took place just as the Sun came up. You normally had been up for 24 or 36 hours getting your experiment ready, you had consumed innumerable cups of strong black coffee, you had suffered innumerable small and large panics over your instruments, and when you came out to the flight line just at dawn you found everybody standing around with their dampened fingers in the breeze trying to decide if it would be safe to launch and, if so, in which direction to lay out the load line. You wanted a gentle breeze blowing down the load line so when you released the balloon it rose majestically from the hold-down arms, moved up over the load line, gently picking up the payloads as it rose. If you didn't get it laid out right or the wind changed direction then you might find yourself and a

companion or two running across a corn field holding a 300-pound payload in your arms to prevent it being dragged by the balloon as it lifted off. A 100-foot tall silvery balloon rising gently all rosy in the light of the rising Sun was a magnificent sight. However, there was one problem with a balloon. When you had trouble you had lots of time to watch it unfold. A wind might come up and slowly wind up your balloon, cutting off the helium supply before you could launch it. A rip might develop on liftoff and you could watch very carefully as the rip went up the side of the balloon, the helium escaped, the balloon collapsed, and you were left with nothing but an acre or two of this plastic spread over the landscape to pick up and carry home.

Frank, with his usual aplomb, managed to entwine his long legs in the load line on one launch and was well on his way to being launched head down when two graduate students held down the load line and extricated our hero, saving him and a considerable portion of the future space science program.

A rocket launch was different. True, you were up all night and you had the same panics with your instruments. In addition, a successful launch, with all the sound and fury and the streak of flame into the sky, is also a magnificent sight. Your troubles with a rocket were quite different, however. Things happen very fast at a rocket launch. There is a flash of light and a roar and your rocket is gone or the tower is demolished. There is none of the slow, genteel failure of a balloon. When a rocket goes bad, it either blows up in the first few seconds of launch or else gets off all right and mercifully comes apart well out of sight.

These rocketeering astronomers and upper atmospheric physicists were similar in many respects to the cosmic ray physicists. They were interested in studying and understanding natural phenomena, they published in similar journals, and they had their movable feast, the Rocket and Satellite Research Panel which met to discuss the significance of their results, plan new rockets, arrange expeditions, and discuss how to get funding for their experiments. They had to spend a lot of time discussing funding because a scientific paper based on sounding rocket research cost the government a good deal more than a paper based on balloon research. They too went on safaris to exotic places such

as White Sands and Fort Churchill, usually trying to make White Sands in July and Churchill in January. Homer Newell, Jack Townsend, John Lindsay, Les Meredith, and John Clark were among those who came to NASA with that background.

There were a few scientists who were not satisfied with the difficulties and frustrations of using either balloons or rockets but were compelled to combine the two. Jim Van Allen formed such a group in Iowa in the early 1950's. If you used an unreliable balloon to carry an unreliable rocket to a high altitude, and if you were very, very lucky you could get your instruments into places instruments had never been before. To do this you assembled balloons and rockets on board a Navy ship and sailed off to the north where you had a small but finite probability of studying the aurora and very low energy cosmic rays. As you would expect, Frank was strongly attracted to such a dubious undertaking. Our hero of the day, now Dr. McDonald, a husband and parent, headed south from Minneapolis, stopped off in Iowa, joined Van Allen's group and headed back north on one of these expeditions. The careful, thorough experimenter, dedicated to understanding every detail of the tools he used, Frank set off a rocket on the deck of the ship thereby convincing himself, and his shipmates, once and for all, that rockets are dangerous and demonstrating that in addition to a penchant for tackling tough problems he also had a pretty hard head, a characteristic he is finding invaluable these days as NASA's Chief Scientist.

The third group of people who came together in the formative days of NASA was funded by and operated under the guidance of the National Advisory Committee for Aeronautics. The NACA was an ancient and honorable institution, having been formed in World War I to enable the United States to catch up with European aeronautical research. This group of people was located at three NACA centers: Langley, Lewis, and Ames. They too were engaged in research but of quite a different nature than the first two I described. The first two groups were interested in understanding natural phenomena. In contrast, this third group studied the behavior of manmade objects, airplanes, rockets, satellites, and reentry vehicles. They used wind tunnels to improve the performance of airplanes and they used airplanes themselves as research tools. They studied the behavior of models of rockets in their wind

tunnels and they used rockets to propel aircraft and rocket models through the atmosphere to understand the phenomena of high speed flight. Sometimes, they used the rockets themselves as research instruments to study the behavior of rockets and develop models to predict their performance. They used rockets to drive nose cones into the atmosphere to perfect heat shields. The first two groups were also interested in balloons, rockets, and airplanes, but strictly for transportation of their instruments. The third group was interested in studying and understanding the behavior of airplanes and rockets. The goal of the third group was to improve the human ability to travel and operate in the atmosphere and space.

These people were mostly bright, young engineers who came to NACA center as young engineers with bachelor's degrees fresh out of college. Many joined NACA directly after World War II, others received their degrees under the GI Bill and joined NACA in the late 1940's and early 1950's. They learned to do their research at the NACA center under the tutelage of a senior research scientist or engineer, rather than at a university under a professor as those in the other two groups did.

They published their own refereed NACA journals. They worked with academic people but tended to use them to carry out center-instituted research programs rather than sponsoring independent research work at universities. They organized their institutions around particular kinds of research, with Langley, the mother center, specializing in conventional aeronautical research, and spinning off the Ames center to do hypersonic research and the Lewis Center to do propulsion research. They worked very closely with the aeronautical departments of the Air Force, Army, and Navy and the aeronautical industry and measured their success by the usefulness of their research to those departments and that industry. They were accustomed to conceiving, designing (and sometimes building), and testing new airplanes, rockets, and spacecraft. Usually the hardware was built by an aerospace company for one of the services. They too went on safaris to exotic places such as the High Speed Flight Test Center at Edwards Air Force Base and Wallops Island. They admired and supported the dedicated professional test pilots, people such as Chuck Yeager with the real "right stuff". Living and working on the shore of the Chesapeake Bay, they liked seafood.

Jesse Mitchell, Ed Cortright, George Low, Abe Silverstein, and Harry Goett were among those who came from NACA and had a substantial impact on the early space science program.

There were some common elements in the ONR and NACA culture. Both groups were extremely proud of themselves and their work. Both groups recognized and appreciated people who did good work and knew their field. Both groups were intensely interested in space. Both recognized its challenge and its potential. Both felt it important that the nation's space effort should be directed by a civilian agency and both ultimately recognized that they needed to work together if the program was to succeed and realize its full potential.

It was to the NACA, in October 1958, that the nation, frightened by Sputnik, gave the responsibility to organize and conduct a National Space Program, but the charter establishing the Space Program placed the objective of the ONR culture—"the expansion of human knowledge of phenomena in the atmosphere and space"—first, and the objective of the NACA culture "the improvement of the usefulness, performance, speed, safety, and efficiency of aeronautical and space vehicles" second. The act added a third objective—"the development and operation of vehicles capable of carrying instruments, equipment, supplies, and living organisms in space". The allocation of NASA's resources among these three objectives has been the subject of continuous and often acrimonious debate for the past quarter century.

People from the NACA centers came to Washington in October, 1958 to form NASA. They were joined shortly thereafter by a group from NRL. The cosmic ray physicists, including Frank, began to drift into the space science program during 1959. Some joined NASA, as Frank and I did, others, such as Kinsey Anderson and Bill Webber, remained in academia, choosing to conduct their research there and influence the NASA Program through service on the Space Science Board and a variety of NASA Advisory Groups.

Very early, consistent with NACA tradition and in recognition of a definite need, NASA established a new spaceflight center, initially the Beltsville

Spaceflight Center and later the Goddard Space Flight Center. NASA also acquired the Jet Propulsion Laboratory as a second space flight center.

Very early, and also consistent with NACA tradition, these new centers were given roles and missions. NASA assigned to GSFC the responsibility for satellites, sounding rockets, manned space flight, and the tracking and communications facilities to support them. To JPL it assigned the responsibilities for the lunar and planetary missions and the deep space communications and tracking net.

Based largely on the recommendations of the Space Science Board, chaired by Lloyd Berkner, what ultimately became a 10-year space science program was established including OAO, OGO, RANGER, SURVEYOR, MARINER, and some 20 or 30 explorers. All of this was laid out and underway by late 1959 and early 1960. I say it ultimately became a 10-year program because it was 1969 before we launched the last Surveyor and the last OAO. When the program was laid out we expected to complete it by 1965, however in those early days things didn't always work out as planned and a program might slip for a year or two.

From the very beginning there was a certain amount of friction between the ONR culture and the NACA culture and a tendency to choose up sides and disagree about something, usually the allocation of resources. In the very early days it was the "NACA Crowd" versus the "NRL Crowd". Occasionally it surfaced as an argument between the scientists and the engineers, and finally with the advent of Apollo it took on its present form "Manned Space Flight" versus "Unmanned Space Flight", which was kind of an ironic twist since most of the leaders of the Manned Space Flight Program such as Robert Gilruth came from PARD, the "Pilotless Aircraft Research Division" at Langley.

I am happy to report both cultures continue to flourish in NASA and both are squared off ready for battle over the space station. The heirs of the ONR culture are concerned about the impact of the space station on the resources available for science. They are not sure there is a real justification for the space station since it is not clear how it will be used for science. Such talk still infuriates those from the NACA culture. To them there is no need for science to justify the space station. The justification for the space station is

in the second and third objectives of NASA's charter. To the heirs of the NACA culture the space station is the next logical step in the development of the human ability to move and operate in the atmosphere and space. It is the next step in a long, honorable process that produced the DC-3, X-15, 747, Apollo, Skylab, and the Shuttle.

The survival of both cultures has been and continues to be one of NASA's great strengths and is one of the principle reasons for the success of the NASA Program. Had the ONR culture prevailed we would probably have had a more pure, more sharply focused, but considerably diminished science program. We would not have had Apollo, Viking, Skylab, or the Shuttle. Had the NACA culture prevailed we would have had Apollo and Skylab but probably not the explorers, IMP's, OSO's, OAO, the space telescope, the Pioneer, the Mariner, Viking, or Voyager. We would have had strong space flight centers but not strong academic groups carrying out their own research program.

Despite, or more accurately because of, the clash and the competition between these two cultures we have had a far better, far richer, more useful, and more successful space program.

All of us here this afternoon owe Frank McDonald a major debt for a quarter century of effort to first instill and then maintain the ONR culture in NASA. He has been a steady, consistent, and powerful advocate for creativity, excellence, and economy in science and for a space science program open to all the world's scientists on the merits of their ideas and their abilities and not on the basis of their political power, their institution, or their national allegiance. The Goddard Space Flight Center, the space science program, NASA, and many of your careers would have been quite different without Frank.

We are all proud to have known and worked with you, Frank. We wish you a happy sixtieth birthday and many more.

BIBLIOGRAPHIC DATA SHEET

1. Report No. NASA CP-2464	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Essays in Space Science		5. Report Date June 1987
		6. Performing Organization Code 665
7. Author(s) R. Ramaty, T. L. Cline, and J. F. Ormes, Editors		8. Performing Organization Report No. 87B0055
9. Performing Organization Name and Address Goddard Space Flight Center Greenbelt, MD 20771		10. Work Unit No.
		11. Contract or Grant No.
		13. Type of Report and Period Covered Conference Publication
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546		
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

This volume consists of the papers presented at a Symposium held at the Goddard Space Flight Center on April 23, 1985 and additional papers written by individuals closely associated with Frank McDonald during various aspects of his career. The papers in Essays in Space Science cover a broad segment of space research and are an acknowledgement of Frank McDonald's personal involvement in many of these efforts. The totality of the papers were chosen so as to sample the scientific areas influenced by him in a significant manner. The papers are arranged into three broad areas: particles and fields of the solar system, cosmic ray astrophysics, and gamma ray, X-ray and infrared astronomies.

17. Key Words (Selected by Author(s))	18. Distribution Statement
Space Science, Astrophysics	Unclassified - Unlimited Subject Category 88

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price*
Unclassified	Unclassified	438	A19

*For sale by the National Technical Information Service, Springfield, Virginia 22161

GSFC 25-44 (10/77)

END

DATE

FILMED

AUG 20 1987